Abstract:
This document identifies and discusses relevant aspects that influence the future sustainability of e-Science Cloud infrastructures. Moreover, the document presents a current market analysis along with an analysis of the VENUS-C User community in terms of behaviour and requirements. VENUS-C takes on-board the after mentioned to help define its position in the Cloud landscape. VENUS-C has demonstrated cloud value-add for the long-tail of science and small businesses; contributing to an early evaluation of the benefits, drivers and challenges addressed that merit further investigation.
### Document Information Summary

<table>
<thead>
<tr>
<th>Deliverable number</th>
<th>D 3.10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deliverable title</td>
<td>Future Sustainability Strategies – final</td>
</tr>
<tr>
<td>Editor</td>
<td>Nadia Nardi/ENG</td>
</tr>
<tr>
<td>Contributing Authors</td>
<td>Nadia Nardi/ENG, Fotis Karayannis/MIC-GR, Stephanie Parker/OGF</td>
</tr>
<tr>
<td>Reviewers</td>
<td>Daniele Lezzi/BSC, Angela Sanger/COSBI</td>
</tr>
<tr>
<td>Work package no.</td>
<td>WP3</td>
</tr>
<tr>
<td>Work package title</td>
<td>Communication and Sustainability</td>
</tr>
<tr>
<td>Work package leader</td>
<td>OGF</td>
</tr>
<tr>
<td>Work package participants</td>
<td>ENG, EMIC, OGF, BSC, UPVLC, KTH, AEGEAN, COSBI, UNEW, CNR, COLB, MIC-GR</td>
</tr>
<tr>
<td>Distribution</td>
<td>Public</td>
</tr>
<tr>
<td>Nature</td>
<td>Report</td>
</tr>
<tr>
<td>Main Target Audiences</td>
<td>VENUS-C partners, research communities, VENUS-C IT developers, European policy makers, public procurement bodies</td>
</tr>
<tr>
<td>Version/Revision</td>
<td>v.final</td>
</tr>
<tr>
<td>Draft/Final</td>
<td>Final</td>
</tr>
<tr>
<td>Keywords</td>
<td>Sustainable scientific Cloud, market analysis, Cloud market, e-Science, usage models, Cloud economy, Business Models</td>
</tr>
</tbody>
</table>

### DISCLAIMER

VENUS-C is a Research Infrastructure Project co-funded by the GÉANT & e-Infrastructure Unit of the Information Society & Media Directorate General of the European Commission.
VENUS-C targets scientific communities, decision makers from SMEs, scientific clusters, incubators and technology parks, as well as European Union policy makers. VENUS-C is supported and evaluated by an Expert International Advisory Committee (EIAC).

This document has been produced with the co-funding of the European Commission. The content of this publication is the sole responsibility of the VENUS-C consortium and cannot be considered to reflect the views of the European Commission.

Grant Agreement Number: 261565

Start Date: 1 June 2010

Duration: 24 Months
## Log Table

<table>
<thead>
<tr>
<th>Issue</th>
<th>Date</th>
<th>Description</th>
<th>Author/Partner</th>
</tr>
</thead>
<tbody>
<tr>
<td>V0.1</td>
<td>29/02/2012</td>
<td>Table of contents</td>
<td>Nadia Nardi/ENG, Andrea Manieri/ENG, Fotis Karayannis/MIC-GR</td>
</tr>
<tr>
<td>V1.0</td>
<td>24/04/2012</td>
<td>First Draft</td>
<td>Nadia Nardi/ENG</td>
</tr>
<tr>
<td>V1.1</td>
<td>03/05/2012</td>
<td>Added feedback/guidance</td>
<td>Andrea Manieri/ENG</td>
</tr>
<tr>
<td>V2.0</td>
<td>08/05/2012</td>
<td>Integration of feedback (draft 2)</td>
<td>Nadia Nardi/ENG</td>
</tr>
<tr>
<td>V2.1</td>
<td>17/05/2012</td>
<td>Added contributions/review</td>
<td>Fotis Karayannis/MIC-GR,</td>
</tr>
<tr>
<td>V2.2</td>
<td>19/05/2012</td>
<td>Integration of feedback (draft 3)</td>
<td>Nadia Nardi/ENG</td>
</tr>
<tr>
<td>V2.3</td>
<td>21/05/2012</td>
<td>Added contributions/review</td>
<td>Fotis Karayannis/MIC-GR,</td>
</tr>
<tr>
<td>V3.0</td>
<td>23/05/2012</td>
<td>Integration of feedback and other modifications (draft 4)</td>
<td>Nadia Nardi/ENG</td>
</tr>
<tr>
<td>V4.0</td>
<td>29/05/2012</td>
<td>Integration of ExComm feedback</td>
<td>Nadia Nardi/ENG</td>
</tr>
<tr>
<td>V4.1</td>
<td>07/06/2012</td>
<td>Added contributions from partners (Stephanie Parker/OGF, Ignacio Blanquer/UPV) (draft 5)</td>
<td>Nadia Nardi/ENG</td>
</tr>
<tr>
<td>V4.1.1</td>
<td>08/05/2012</td>
<td>Review by Reviewers</td>
<td>Daniele Lezzi/BSC, Angela Sanger/COSBI</td>
</tr>
<tr>
<td>V4.2</td>
<td>13/06/2012</td>
<td>Addressed reviewer comments (draft 6)</td>
<td>Nadia Nardi/ENG</td>
</tr>
<tr>
<td>V4.3</td>
<td>14/06/2012</td>
<td>Overall revision and submission to PMB</td>
<td>Andrea Manieri/ENG, Nadia Nardi/ENG</td>
</tr>
<tr>
<td>V4.4</td>
<td>22/06/2012</td>
<td>Addressed comments by Silvana Muscella/OGF and Goetz Brasche/EMIC, and integrated additional contribution by Stephanie Parker/OGF (draft 7)</td>
<td>Nadia Nardi/ENG</td>
</tr>
<tr>
<td>V4.5</td>
<td>25/06/2012</td>
<td>Addressed comments by Furio Barzon/COLB and Goetz Brasche/EMIC and finally approved by PMB</td>
<td>Andrea Manieri/ENG</td>
</tr>
<tr>
<td>v.final</td>
<td>25/6/2012</td>
<td>Quality Assurance check</td>
<td>Fotis Karayannis/MIC-GR</td>
</tr>
</tbody>
</table>

## Document Approval

<table>
<thead>
<tr>
<th>Approval by</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Project Management Board</td>
<td>25/06/2012</td>
</tr>
</tbody>
</table>
EXECUTIVE SUMMARY

VENUS-C is the first European distributed computing infrastructure to adopt a User-centric approach to Cloud Computing, putting the needs of end-User communities of researchers and small businesses at the forefront of development, and providing scalable and interoperable cloud resources that combine both open source and commercial solutions to offer the best of both worlds. The VENUS-C vision is to support primarily the ‘long tail of science’\(^1\), that is, the vast majority of researchers who may have never had access to supercomputer networks and have relied on desktop resources. The aim has been to empower them in a number of different ways, enabling them not only to do better science by accelerating discovery but also new science they could not have done before. Research disciplines now using the cloud are as diverse as astrophysics, biodiversity, civil protection and emergencies, geosciences, mathematics, healthcare and medical research, cellular, genetic and molecular biology, and social media assessment.

The VENUS-C User-centric approach and the Cloud Computing services it has successfully provided to our wide User community is an important example of how to create services tailored to scientists belonging to the above mentioned ‘long tail of science’. We are well aware that getting access to what a scientist needs, for example computational power and storage resources can prove very difficult for those scientists and SME’s who have limited funds and facilities. VENUS-C, from its inception, has used this assumption to position itself among the rest of the DCI projects, being complementary and taking care of the needs that other DCI projects don’t address. VENUS-C’s proposed architecture and the exposed APIs are thus suitable for many such disciplines or applications of the long tail and not only. A representative User community sample was involved in the project through the 7 scenarios, the 15 pilots and the 5 experiments.

This sample has covered a number of small businesses, developing new services without the need for large up-front capital investment. Chief among the success stories is Green Prefab, a start-up company for architects and civil engineers, which hosts a new innovation cluster stemming from VENUS-C, HUB-Engineering (http://www.hub-e.com/), taking as its starting point two complementary services, one on the structural analysis of buildings provided by the Universidad Politécnica de Valencia (VENUS-C partner), the other, called EnergyPlus, led by the Danish Royal Academy (VENUS-C pilot). Molplex Ltd, a VENUS-C pilot and UK start-up, is now able to provide cloud-based drug discovery on demand to its growing market. A Swiss small business, DFRC (also VENUS-C pilot) is benefitting from the infrastructure to scale their platform and enable it to support future growth in terms of vessels monitored in real-time and usability by operators. More broadly, we found that Cloud Computing offers important opportunities for collaboration between research-focused SMEs and academia. Austria-headquartered RISC-Software, for example, is using its experience in VENUS-C to assist technology transfer between research and industry. Specifically, it is working with the University of Malaga in Spain and the Johannes

---

\(^1\) [http://en.wikipedia.org/wiki/Long_Tail](http://en.wikipedia.org/wiki/Long_Tail)
Kepler University in Austria to harness the collaborative power of the Cloud Computing and benefit from the open approach adopted in the VENUS-C project.

The main technical goals of the VENUS-C project have been to design, test and deploy an industry-quality cloud infrastructure tailored to the requirements of the reference User community featured above. Technical work has focused on enhancing existing and developing middleware components for secure job and data management, including programming models, data management proxies, application security, monitoring, accounting and billing, as well as cloud network optimisation. The VENUS-C Architecture Report describes the blueprint for the platform based on which a reference implementation has been done applying a rapid prototyping approach with interactive releases and User validation. The experiences gained in adapting applications for the cloud and in performing core technical work are being shared with the wider community in a dedicated in-house publication along with complementary publications mainly dedicated to the benefits gained by the VENUS-C User community in real-world settings.

Our sustainability study and User validation have revealed that different factors come into play when it comes to evaluating migration to the cloud and defining the conditions under which the tangible and intangible assets accrued can effectively become sustainable. It became clear that a single cloud benefit will not suffice as a driver but rather the interplay of several factors, including different expectations, experiences and skill sets. For example, we have revealed that through the findings of our sustainability questionnaire and current literature that cost savings are not at the forefront of researchers’ mind when deciding to use the cloud, rather they are driven by better performance or a new capability. A UK report, which explores the cost of Cloud Computing for research concludes that it is likely that the increased use of Cloud Computing will lead to more and better science but not necessarily to lower costs. Our study is aligned with these findings, particularly in terms of the different tangible benefits that have been gained by the User community. Another important VENUS-C finding is that technical support is a vital ingredient also for Cloud Computing. For this reason a special helpdesk has been set up to provide close to real-time practical support, complemented by a dedicated eTraining environment and physical training events for the 7 scenarios and 15 pilots which joined VENUS-C in summer 2011.

With regard to tangible and intangible assets, the main VENUS-C outputs, released under open source licences and exploitable to ensure sustainability, are shown in the list below. Near future plans include drawing on the value proposition around the exploitable assets and building a community around the open-source components with the aim of fostering further exploitation by research communities and small businesses.

**Tangible assets**

**Public release of open-source components**

---

2 Cost for cloud computing for research, Final Report to EPSRC and JISC by Curtis & Cartwright, UK. http://www.jisc.ac.uk/media/documents/programmes/research_infrastructure/costcloudresearch.pdf
- Generic Worker (PMES (Process Management Enactment Services) and programming framework)
- COMPSs (PMES and programming framework)
- CDMI Proxy
- Accounting and billing service
- Traffic Redundancy Elimination (TRE)
- OVF4ONE
- Seven Scenario Applications (Wildfire, RainyCloud, Bio Cloud Tools, Architrave, Green Prefab HUB-E, COSBI, QSAR)

**Intangible assets**

- eTraining environment with forthcoming packages for autonomous end-User support
- eTraining environment with educational services for the wider community
- Project sponsor credibility and reputation in the marketplace
- Commitment to relevant standards, including high-level cloud standards and established protocols, like VENUS-C CDMI and OCCI implementation in the EGI Federated Cloud Test Bed
- In-house and peer-review publications on the main assets accrued
- Visibility in the European landscape, including the Digital Agenda for Europe

**General advancement of knowledge (experience and know-how)**

1. Open Nebula for eScience
2. Azure for eScience
3. Porting of an application to the Cloud
4. eScience Computing environment
5. Deep insights into Cloud Computing
6. Access to VENUS-C User network (possible future business/collaborations)
7. Access to experiences of all VENUS-C User communities (pilots + scenarios + experiments)
8. Training Material/Documentation

Our sustainability questionnaire findings from the VENUS-C User community and related research analysis lead us to the affirm that there is a true potential for Scientific Clouds to be sustainable, in the near term in the form of hybrid Clouds, composed of both private/community clouds and public Clouds offered by commercial cloud providers through a procurement process. The hybrid Cloud could therefore be considered an attractive model that brings together the best of the two worlds, i.e. the tailored-feel of a private/community Cloud and the proven economies of scale of a public Cloud. The private/community Cloud will be serving the needs of the big and well-organised communities with complex requirements, while the commercial public cloud resources will serve the long tail of scientists and possibly some of the cloudbursting needs of the first.

Thanks to the open collaboration among DCIs and the wide dissemination between each other, and of course the great benefits that Cloud Computing has for scientific research, a new hybrid environment looks promising and has the potential to inspire and create innovation.

As a result, it is clear that the adoption of Cloud and Virtualisation technologies and services is a win-win situation for the scientific community, as it releases benefits for both research institutions, industry as

---

3 Hostingtecnews.com published an article on the Wildfire application describing it as “new service has potential for many other regions across the globe affected by devastating fires.” http://www.hostingtecnews.com/european-venus-c-project-empowering-emergency-crews-cloud
well as other research stakeholders (e.g. the European Commission that safeguards its investments and in parallel seeks innovation) and SME’s by taking advantage of economies of scale, ease of entry and higher productivity levels.

Acknowledgment
The VENUS-C project team wants to especially thank the EIAC (experts) who have played an important role in helping to define the technical value added of VENUS-C.
TABLE OF CONTENTS

EXECUTIVE SUMMARY........................................................................................................... 5

1. INTRODUCTION.................................................................................................................. 15
   1.1 Scope of the Document.................................................................................................. 15
   1.2 Target Audiences........................................................................................................ 15
   1.3 Structure of the Document........................................................................................ 15
   1.4 Terms and Definitions................................................................................................ 17

2. MARKET ANALYSIS AND STRATEGIES FOR e-SCIENCE CLOUDS.......................... 18
   2.1 Public Cloud Market Analysis and Trends................................................................. 18
       The Landscape.............................................................................................................. 19
       European Providers..................................................................................................... 22
       The European View.................................................................................................... 22
   2.2 e-Science Clouds Market Analysis............................................................................ 24

3. SCIENTIFIC CLOUD ANALYSIS....................................................................................... 25
   3.1 VENUS-C Sustainability Questionnaire....................................................................... 25
   3.2 Scientific Cloud SWOT Analysis............................................................................... 25
   3.3 Compute Models and Scientific Behaviour.............................................................. 27
       Local Cluster Systems versus Cloud Computing...................................................... 28
       Our User Community............................................................................................... 28
       Findings..................................................................................................................... 29
       Resource Usage Models........................................................................................... 35
       Usage Behaviour and Cost Effectiveness................................................................... 36
       Load Profiles.............................................................................................................. 37
       Actual Usage Patterns............................................................................................... 38
       “Home is where the Cloud is”................................................................................ 40
   3.4 Gap Analysis for Adoption of Cloud Computing by e-Science............................... 40
       An Overview................................................................................................................ 40
       Non-Technical Gap - Legislation, Government & Policies......................................... 43

4. SUSTAINABILITY ANALYSIS.............................................................................................. 43
4.1 Strategies for e-Science Clouds ................................................................. 43
  Private/Community Federated Cloud for e-Science ...................................... 44
  Sustainability Approaches for e-Science Clouds ......................................... 46
  Summary of considered Business Models ..................................................... 47
4.2 Sample of VENUS-C Business Models ..................................................... 49
4.3 Technical Elements .................................................................................... 52
4.4 Organisational Elements ............................................................................ 53
4.5 Financial Elements ..................................................................................... 54
4.6 Policy Elements ......................................................................................... 55
4.7 Legal Elements ............................................................................................ 55
4.8 Other Elements ........................................................................................... 56
5. VENUS-C SUSTAINABILITY AND EXPLOITATION STRATEGY .................. 57
  5.1 VENUS-C Sustainability Strategy and Exploitable Assets ......................... 57
     Tangible Assets ............................................................................................ 57
     Intangible Assets .......................................................................................... 59
  5.2 Partner and Pilot Final Comments on VENUS-C ....................................... 61
  5.3 Partners Sustainability and Exploitation Plans .......................................... 65
     ENGINEERING - INGEGNERIA INFORMATICA SPA .................................. 65
     OGF.eeig ...................................................................................................... 66
     COLB - COLLABORATORIO ....................................................................... 68
     UNEW - UNIVERSITY OF NEWCASTLE ..................................................... 69
     UPVLC - UNIVERSIDAD POLITECNICA DE VALENCIA ............................ 70
     KTH - KUNGLIGA TEKNISKA HOEGSKOLAN ........................................ 71
     COSBI - UNIVERSITY OF TRENTO CENTRE FOR COMPUTATIONAL AND SYSTEMS BIOLOGY ................................................................. 71
     BSC - BARCELONA SUPERCOMPUTING CENTER ....................................... 72
     CNR - CONSIGLIO NAZIONALE DELLE RICERCHE .................................... 73
     Microsoft (MRL- MICROSOFT RESEARCH LIMITED, EMIC- EUROPÆISCHES MICROSOFT INNOVATIONS CENTER GMBH, MICROSOFT INNOVATION CENTER GREECE) ........................................................................................................ 75
IIT - TECHNION - ISRAEL INSTITUTE OF TECHNOLOGY ........................................... 76
AEGEAN – UNIVERSITY OF AEGEAN .................................................................. 77

6. CONCLUSIONS .................................................................................................. 78
7. REFERENCES & RESOURCES ........................................................................... 80
LIST OF FIGURES
Figure 1 – Hype Cycle for Cloud Computing, 2011.................................................................19
Figure 2 – Magic Quadrant for Public Cloud Infrastructure as a Service ......................20
Figure 3 – Workload Distribution (2010-2015) ........................................................................22
Figure 4 – Questionnaire Answer – How do you manage the data for your application?..........................................................................................32
Figure 5 – Questionnaire Answer – Do Virtual Machines Effect Your Application? ....33
Figure 6 – Questionnaire Answer – Top 5 Cloud Computing features which are most attractive ........................................................................................................34
Figure 7 – Questionnaire Answer – Whether the Cloud is technically sound ..............35
Figure 8 – Cumulative Normalised CPU time - Life Sciences on EGI infrastructure (June 2011 to May 2012).. ............................................................36
Figure 9 – Choosing the Most Cost-Efficient Computing Option..................................37
Figure 10 – Daily CPU consumption (CPU per hours) of the different scenarios ....39
Figure 11 – The Hybrid Model .................................................................................................48
Figure 12 – Cloud Brokering Model.......................................................................................48
Figure 13 – Individual scientist with direct access to public cloud model ..........49
Figure 14 – Luxrender .............................................................................................................51
Figure 15 – EnergyPlus ............................................................................................................51
Figure 16 – Architrave .............................................................................................................52
Figure 17 – VENUS-C Channel snapshot of OS downloadable components ........58
Figure 18 – VENUS-C Channel snapshot of OS downloadable components - Internal page ........................................................................................................58
Figure 19 – Questionnaire Answer – Rating of Elasticity in regards to Job Execution ..............................................................84
Figure 20 – Questionnaire Answer – Rating of Elasticity in regards to Data Management .............................................................................................................85
Figure 21 – Questionnaire Answer – Rating of Reliability in regards to Data Management .............................................................................................................85
Figure 22 – Questionnaire Answer – Rating of Ease of Use in regards to Accounting and Billing ..........................................................................................86
Figure 23 – Questionnaire Answer – Rating of Flexibility in regards to Accounting and Billing .............................................................................................................86
Figure 24 – Questionnaire Answer – Ranking of experience with workflow ............87
Figure 25 – Questionnaire Answer – Ranking of experience with Cloud technologies before the grant .............................................................................................................89
Figure 26 – Questionnaire Answer – Ranking of best Cloud model for your application in the near future ........................................................................94
Figure 27 – Questionnaire Answer – Whether the Cloud service would be recommended to a friend/colleague

Figure 28 – Cumulative Normalised CPU time of the Astrophysics

Figure 29 – Cumulative Normalised CPU time of the Earth Sciences

Figure 30 – Cumulative Normalised CPU time of the Computer Science and Mathematics

Figure 31 – Cumulative Normalised CPU time of the Fusion

Figure 32 – Cumulative Normalised CPU time of the High-Energy Physics

Figure 33 – Cumulative Normalised CPU time of the Computational Chemistry
LIST OF TABLES
Table 1 – Summary of User requirements scoring and mapping to VENUS-C components
Table 2 – Questionnaire answers to run time and frequency
Table 3 – Cloud Computing gaps (source: Advances in Clouds - Research in Future Cloud Computing)
Table 4 – VENUS-C Tangible Assets
Table 5 – VENUS-C Intangible Assets
Table 6 – Exploitable assets mapped to partners
Table 7 – List of Questionnaire Respondents
1. INTRODUCTION

This analysis explores sustainability perspectives of scientific Clouds. It has been conducted to develop a clear representation of the current and potential benefits of the Cloud (considering the different deployment models) for scientific applications. The main objective of this document is thus to formulate a sustainability strategy for e-Science Cloud-based infrastructures.

In detail, an analysis of the following aspects related to Cloud computing was performed: current business models, costs and return of investments, SWOT analysis, gap analysis, the wide literature from Analysts (Gartner, etc.), and other international experiences, also building on past ESFRI and e-IRG recommendations. Furthermore, in investigating sustainable Cloud models for e-Science different related elements were considered being: legal, policy, financial, organizational, technical, barriers due to data privacy and confidentiality, and identity management.

1.1 Scope of the Document

The scope of this document is to analyse the sustainability of a scientific Cloud, as proposed by the VENUS-C project and enabled by the VENUS-C platform. The document has taken in account the current studies on Cloud Computing for eScience and builds on the initial findings and analysis performed in Deliverable 3.9 (Future Sustainability Strategies – interim), along with D3.7, named Potential Legal Issues. Through the analysis and comparison of current Cloud models available, along with the analysis of information gathered from the VENUS-C User community, the document reports the suitability of Clouds in scientific contexts.

1.2 Target Audiences

This document targets:

- VENUS-C partners to communicate the opportunities and potential profitability of the VENUS-C business model.
- Research communities, who intend to use scientific Cloud infrastructures to help them evaluate the benefits and the costs of such infrastructures.
- Public procurement bodies dealing with research budgets.
- European policy-makers, who need to understand the potential of Scientific Cloud and related Cloud API’s for e-Science such as the VENUS-C ones.
- Individual researchers, Cloud providers, data centre owners, to appreciate the benefits of the VENUS-C results.

1.3 Structure of the Document

The document is structured in six main chapters, namely this introduction and five others, which follow the analysis process introduced above:

- The second chapter is dedicated to a market analysis and strategies for e-Science Clouds. You will find information on the public Cloud market and trends.
- In the third chapter, you will find a continuation of the analysis specific to the scientific Cloud. Here you find a description on the sustainability questionnaire distributed to the entire VENUS-C
User community. The analysis of the questionnaire is found throughout the document in the pertinent sections, and the full questionnaire and analysis is found in Annex A. Furthermore, a scientific Cloud SWOT analysis is presented. Sections dedicated to usage models and scientific behaviour in the Cloud are then described and analysed. Finally, a summary gap analysis for the adoption of Cloud Computing for e-Science is detailed.

- In the **fourth chapter**, you will find information on strategies for e-Science clouds, where the following elements are presented: private/community federated Cloud for e-Science, use of public clouds for e-Science, a proposed sustainability approach for e-Science clouds – the hybrid approach, and finally a summary of the possible business models. Furthermore, the following elements will be explored in regards to the sustainability analysis: technical, organisational, financial, policy, legal, and other relevant elements.

- The **fifth chapter** is dedicated to the VENUS-C sustainability and exploitation strategy. Here, the exploitable assets are described. The sustainability and exploitation plans of each of the VENUS-C partners are presented. They include a description of the principal VENUS-c components they will exploit, what the VENUS-C selling factors are according to them, and information on their principal customer base and network.

- In the **sixth chapter** you will find the conclusion, which will wrap-up the entire document, bringing together all the key aspects of what sustainability means for the scientific community and Cloud computing.
1.4 Terms and Definitions

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>API</td>
<td>Application Programming Interface</td>
</tr>
<tr>
<td>BT</td>
<td>British Telecommunications plc</td>
</tr>
<tr>
<td>CDMI</td>
<td>Cloud Data Management Interface</td>
</tr>
<tr>
<td>COMPSs</td>
<td>Comp Superscalar programming framework</td>
</tr>
<tr>
<td>EGI</td>
<td>European Grid Infrastructure</td>
</tr>
<tr>
<td>e-Infrastructure</td>
<td>An open system that supports flexible cooperation and optimal use of all electronically available resources</td>
</tr>
<tr>
<td>e-IRG</td>
<td>e-Infrastructure Reflection Group</td>
</tr>
<tr>
<td>e-Science</td>
<td>Computationally intensive science deploys highly distributed network environments, or uses immense data sets. The term was created by John Taylor, the Director General of the United Kingdom's Office of Science and Technology in 1999.</td>
</tr>
<tr>
<td>ESFRI</td>
<td>ESFRI, the European Strategy Forum on Research Infrastructures, is a strategic instrument to develop the scientific integration of Europe and to strengthen its international outreach.</td>
</tr>
<tr>
<td>GW</td>
<td>Generic Worker programming model</td>
</tr>
<tr>
<td>HPC</td>
<td>High Performance Computing</td>
</tr>
<tr>
<td>IaaS</td>
<td>Infrastructure as a Service</td>
</tr>
<tr>
<td>PaaS</td>
<td>Platform as a Service</td>
</tr>
<tr>
<td>PMES</td>
<td>Process Management Enactment Services</td>
</tr>
<tr>
<td>PRACE</td>
<td>Partnership for Advanced Computing in Europe initiative aims at developing a European top Supercomputer</td>
</tr>
<tr>
<td>SaaS</td>
<td>Software as a Service</td>
</tr>
<tr>
<td>SME</td>
<td>Small, Medium-sized Enterprise</td>
</tr>
<tr>
<td>STS</td>
<td>Security Token System</td>
</tr>
<tr>
<td>SWOT</td>
<td>SWOT analysis is a strategic planning method used to evaluate the Strengths, Weaknesses, Opportunities, and Threats involved in a project or in a business venture.</td>
</tr>
<tr>
<td>TRE</td>
<td>Traffic Redundancy Elimination</td>
</tr>
</tbody>
</table>

Table 1 – Terminology
2. MARKET ANALYSIS AND STRATEGIES FOR e–SCIENCE CLOUDS

The chapter provides updates from our previous analysis done in Deliverable 3.9– Future Sustainability Strategies – draft, where a report on the following was given: the target market applicable for the e–Science, the different actors involved, the procurement process and best-practices, the usage schemas for local, national or European resource providers and the potential benefits of e-Science Clouds, as enabled by the VENUS-C platform.

This chapter begins with a market analysis on Public Clouds and the related trends, moving on with an analysis of the e-Science Cloud market. The market analysis will continue into chapter 3, where a SWOT analysis of scientific Clouds is presented, along with a summary of scientific behaviour in the Cloud and a gap analysis, also including a comparison to other approaches.

2.1 Public Cloud Market Analysis and Trends

According to the official NIST definition, "Cloud Computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction."4

In taking a closer look at public Cloud Computing, its weaknesses include: data management, trust, control and guarantees on performance: In Gartner’s latest study Hype Cycle for Cloud Computing, it is explained that “the leading concerns expressed by those considering the use of public Cloud Computing services are security, privacy, compliance and regulatory approval. These concerns stem from the fact that public Cloud Computing services are provided via an open Internet-based model to the general public, and that the workloads and data of all consumers run in a common, shared environment."5

Below (Figure 1), according to Gartner6, currently public Cloud Computing, whose years to mainstream adoption is shown as 2-5 years (time required to reach the Plateau of Productivity), is currently at the end of the Peak Inflated Expectations phase, when overenthusiasm kicks-in.

---

Cloud Computing has been actively adopted by many consumer-facing firms who deliver products in large volumes. Commercial service providers are expanding their available Cloud offerings to include the entire stack of IT hardware and software infrastructure, middleware platforms, application system components, software services and turn-key applications. The private sector has taken advantage of these technologies to improve resource utilisation, increase service responsiveness and accrue meaningful benefits in efficiency, agility and innovation. Similarly, for research organisations, Cloud Computing holds significant potential to deliver public value by increasing operational efficiency and responding faster to constituent needs.⁷

The Landscape
There are many Cloud active players, some of which Gartner has chosen to highlight in its latest publication of the *Magic Quadrant for Public Cloud Infrastructure as a Service (2011)*⁸ which provides an overview of who, according to them, the key players are. The inclusion and exclusion criteria in considering which vendors to include are: (1) They must sell it as a stand-alone service, without the requirement to bundle it with managed hosting, application development, application maintenance or other outsourcing. (2) It must be enterprise class, offering 24/7 customer support (including phone

---

⁸ Gartner, Magic Quadrant for Public Cloud Infrastructure as a Service, December 2011 ID:G00226914
support), SLAs and the ability to scale an application beyond the capacity of a single physical server. (3) It must be offered in a minimum of two data centres, located in different cities. (4) They must have significant market presence in public cloud Iaas services (by the definition used in this Magic Quadrant), as demonstrated by more than 1,000 VMs in production use in their public cloud, or more than $10 million in 2010 cloud compute Iaas revenue. Below is the quadrant:

Figure 2 – Magic Quadrant for Public Cloud Infrastructure as a Service
Source: Gartner (December 2011)

In the top right-hand quadrant are the “leaders”, rather those that according to Gartner “... have distinguished themselves by offering an excellent service and having an ambitious future road map. They usually serve a broad range of use cases well, although they do not excel in all areas, and they are not necessarily the best providers for a specific need. They have a track record of successful delivery, along with many referenced customers.” The leaders are:

1. Amazon Web Services, who remains the dominant player in public Cloud Computing worldwide, and has cut prices 19 times in just the past six years.
2. CSC, a large, traditional IT outsourcer with a broad range of data centre outsourcing capabilities. Their cloud Iaas architecture called CloudCompute is available in multiple data centres in the U.S., as well as in the U.K., Denmark, Luxembourg and Australia.

3. Verizon/Terremark, who offers a single-tenant environment reportedly offering the level of security many large enterprises and government agencies require.

4. CenturyLink/Savvis, who boasts a network of 48 international data centres with nearly 2 million square feet of floor space.

5. Bluelock, a company, which connects Users’ VMware data centres with its public Cloud, and focuses largely on small and midsize companies.

Placed in the “visionary” quadrant, as the vendors whom ranked less than the “leaders” in the ability to execute are:

1. Rackspace, the co-creator of the OpenStack open source Cloud OS is positioned to remain a leader. In fact, the OpenStack community just shipped its fifth release. They provide traditionally managed hosting as well as public Cloud Platform as a Service (PaaS), and hybrid Cloud services that blend the two technologies.

2. Joyent, whom Dell has chosen to power its Cloud and whose major industry focus in 2012 is on big data and HPC.

The latest round of Cloud Computing forecasts released by Cisco, Deloitte, IDC, Forrester, Gartner, The 451 Group and others show how rapidly Cloud Computing’s adoption in enterprises is happening.

The demand for both private and public Cloud Computing is up, but the degree to which it is expected to grow is significant. IDC’s (International Data Corporation) "Worldwide and Regional Public IT Cloud Services 2011-2015 Forecast” report\textsuperscript{10} shows spending on public IT Cloud services will reach $72.9bn in 2015, a compound annual growth rate of 28%. Moreover, in 2015, public Cloud services will account for 46% of net new growth in overall IT spending in five key product categories – applications, application development and deployment, systems infrastructure software, basic storage, and servers. IDC also predicts that Cloud services are a critical component in a much larger transformation that IDC expects will drive IT industry growth for the next 25 years.

Cisco predicts that Global Cloud IP traffic will increase twelvefold over the next 5 years, accounting for more than one-third (34%) of total data centre traffic by 2015.\textsuperscript{11} The following graphic is from the analysis:

\textsuperscript{10} http://www.idc.com/getdoc.jsp?containerId=prUS22897311
\textsuperscript{11} Cisco Global Cloud Index: Forecast and Methodology, 2010–2015.
European Providers

In a recent report, *Advances in Clouds - Research in Future Cloud Computing*\(^\text{12}\) the European ICT market is described as having its own particular structure, with much smaller and diversified players, and a general focus towards B2B - the European telecommunication industry is a major exception to these cases. This particular set-up demands much more effort on integration, federation and interoperation to create a European wide cloud ecosystem that incorporates and exploits this diversification for richer service provisioning.

**The European View**

IDC forecasts that the total market will grow from €3.4 billion in 2010 to €14.5 billion in 2015. There will be strong growth in all the sub-markets, including Cloud applications, Cloud platforms, Cloud infrastructure, Cloud servers, and Cloud storage. In addition, in many software and hardware markets, a large share of the market’s growth will be due to Cloud services. Directly or indirectly, Cloud services are therefore impacting very large parts of the overall European IT market.

---

The study specific to Europe, “The Cloud Dividend: Part One. The economic benefits of Cloud Computing to business and the wider EMEA economy”\textsuperscript{13} concludes that the European economy and the euro will cope with the strains from the economic crisis by economic growth in the stronger European economies. The study goes on to assert that Cloud Computing is an important driver to help Europe achieve this, as it helps re-establish the competitiveness of Europe’s international trading position, hence boosting export growth. Secondly, Europe’s economic recovery will be boosted by business investment in Cloud Computing. Not only is Cloud Computing an important issue from the micro perspective of boosting the efficiency of individual companies’ IT investment and, hence, general corporate productivity, but also that, especially in the present uncertain economic climate, it will also be a critical macroeconomic factor that is crucial for boosting Europe’s economic growth.

Neelie Kroes, Vice-President of the European Commission and responsible for the Digital Agenda gave a great speech recently at the European Internet Foundation event on Cloud Computing, where she clearly promotes the amazing potential of the Cloud and strongly expresses her point that we need to make sure we are taking full advantage, and go for a strategic and unified approach. Kroes explains that information and communications technology are huge, representing half our productivity growth. According to one new study, by 2016, the EU Internet economy could be over 800 billion euros: over 5% of GDP. The Cloud takes that further. In the five largest Member States alone, over five years, the Cloud could be worth 2000 euros for each and every citizen and create a million new jobs too. Kroes describes what the Cloud can offer saying that our businesses, especially the small businesses and start-ups, can benefit from flexible, cheap and tailored solutions. Without expensive set-up and capital costs: as long, of course, as you don’t have to get involved in extended legal discussions every time. Kroes goes on to explain that it’s not just businesses that could benefit. In the UK, the government expects to save 20% of annual IT expenditure through harmonising the software for e-Government. Scientists and researchers, too, increasingly need to process and share data flexibly – and the Cloud gives them a new tool to do this. Moreover, she touches on the darker side, being the barriers related to enforceability, lock-in, and data protection and portability and expresses the need to remove at best the obstacles for those trying to get into the Cloud, creating a common approach when it comes to issues like standards, security, and lock-in.

A European Commission report titled Innovation Union Competitiveness report 2011\textsuperscript{14} explains that “in 2008, the EU had approximately 1.8 million researchers (FTE): about 690,000 in the Private Sector; 610,000 in the Higher Education System; and 190,000 in the Public Sector. Only half of the researchers in the EU work in the private sector, where research is more closely linked to innovation. In other countries such as the United States, almost four out of five researchers work in the private sector.” The adoption


of Cloud Computing could help create opportunities to narrow the gap in Europe, as Cloud Computing offers great benefits to single scientists, small businesses’ and start-ups.

2.2 e-Science Clouds Market Analysis

At the heart of modern science and technology is data generation and analysis using computational methods. The name we have given to support data federation and collaboration in the areas of analysis, data mining, data visualization and exploration and for scholarly communication and dissemination is e-Science. As described in an article published for the website www.hpcinthecloud.com, “The most important advantage behind the concept of Cloud Computing for scientific experiments is that the average scientist is capable of accessing many types of resources – including data sets and services- without having to buy or configure the whole infrastructure. This is a fundamental need for scientists and scientific applications.”

Currently, the Cloud services for scientists are in a very primary stage. As interest continues to rise, therefore demand for new tools and services and more support will increase and the market will respond accordingly with new offerings to meet the needs.

The market is beginning to take action to creating dedicated services to meet the needs of scientists by leveraging the benefits of Cloud Computing. British Telecommunications plc (BT), one of the world’s leading providers of communications services and solutions, announced on 25 April 2012 it would launch a new Cloud service (BT for Life Sciences R&D), being the first Cloud service designed to enable collaboration within the life sciences industry for increased R&D productivity. BT partnered with Accelrys (a scientific enterprise R&D software and services company). The aim of the service is to improve collaboration, help reduce the costs of research and development (R&D) involved with bringing new drugs to market, and accelerate time to market. The new service is designed to allow customers to comply with the industry’s stringent security, regulatory and compliance requirements in a way that is suitable for many regulated applications a company may wish to deploy.

Our sustainability questionnaire, presented in the next chapter, confirms the scientific community is looking forward to opportunities similar to the one described above. One respondent commented:

“Clouds can accelerate science in many domains [...] It will be great if either Microsoft (as they did in VENUS-C) or other commercial providers will understand this and they will establish special agreement with National Research Centres operating scientific services for domain-specific scientists.”

16 http://www.btplc.com/
18 http://accelrys.com/
3. SCIENTIFIC CLOUD ANALYSIS

This chapter covers the VENUS-C sustainability questionnaire, a SWOT analysis of Scientific Clouds with dedicated sections to usage models and scientific behaviour. Furthermore, a summary gap analysis for the adoption of Cloud Computing for e-Science is detailed.

3.1 VENUS-C Sustainability Questionnaire

The sustainability questionnaire was distributed to our VENUS-C User community, being our scenarios, pilots, and experiments covering several disciplines (Medicine, Civil Engineering, Molecular Cellular & Genetic Biology, Chemistry, Biodiversity & Biology, Maths and Mechanical Engineering, Information and Communication Technology, Earth Sciences, Civil Protection, and Physics). The questionnaire investigated on and explored how the Users feel about Clouds and about VENUS-C. The feedback collected covers aspects of usage behaviour (performance and usability), technical aspects (assessment of the VENUS-C service), financial/funding type questions, and overall-type questions. An analysis of the questionnaire’s valuable answers was performed, and was also utilized to perform the SWOT analysis specific to Clouds for the scientific community.

See Annex A for a list of the respondents, the original version of the questionnaire and a question-by-question analysis. The feedback and detailed answers we received are interesting and were extremely valuable to our analysis; you’ll find more than just graphs representing numbers, but also answers to important and revealing open-ended questions, chosen with value-added purpose.

The questionnaire was created in an online version and can be viewed here: https://docs.google.com/spreadsheet/viewform?formkey=dHVxQ2Y4UVl6bjE0UGhB53RIT3FqM1E6MQ.

Also refer to D5.8 - Report on support scenario for other details.

Findings of the questionnaire are also represented throughout the document in the relevant sections.

3.2 Scientific Cloud SWOT Analysis

<table>
<thead>
<tr>
<th>Strengths</th>
<th>Weaknesses</th>
</tr>
</thead>
<tbody>
<tr>
<td>✓ On-demand access to resources</td>
<td>✓ Could require significant initial effort and skills to port applications (especially some emerging programming models)</td>
</tr>
<tr>
<td>✓ Enable centralized control over User</td>
<td>✓ Lack of tools and documentation to aid scientists to manage and operate virtualized resources (create manage images, manage jobs and data)</td>
</tr>
<tr>
<td>environment</td>
<td>✓ Limited choice of tools available for workflow and data management</td>
</tr>
<tr>
<td>✓ Flexible resource management</td>
<td>✓ Unique security issues</td>
</tr>
<tr>
<td>✓ No initial significant administration overhead</td>
<td>✓ Data confidentiality/protection issues</td>
</tr>
<tr>
<td>✓ Rapid elasticity (allocated CPU power needed)</td>
<td>✓ Data jurisdiction issues</td>
</tr>
<tr>
<td>✓ Suitable for research loads (peaks, not</td>
<td>✓ Limited possibility for tailored contracts</td>
</tr>
<tr>
<td>continuity)</td>
<td>✓ Lack of national/international policies and laws</td>
</tr>
<tr>
<td>✓ Efficiency in service provisioning</td>
<td></td>
</tr>
<tr>
<td>✓ Easy to use</td>
<td></td>
</tr>
<tr>
<td>✓ Breaking down barriers to collaboration and</td>
<td></td>
</tr>
<tr>
<td>increasing synergetic exploitation of data and</td>
<td></td>
</tr>
<tr>
<td>models</td>
<td></td>
</tr>
<tr>
<td>✓ Significant potential to deliver public value by</td>
<td></td>
</tr>
</tbody>
</table>
increasing operational efficiency and responding faster to constituent needs
✓ Capability of accessing many types of resources without having to buy or configure the whole infrastructure
✓ Allows entities to concentrate economic resources on needs that are not related to IT requirements

*Below, benefits collected from the sustainability questionnaire
✓ Real-time processing
✓ Scalability (increase or decrease resources as needed) / pay on demand / as you go
✓ Cost saving on bigger, faster, newer hardware
✓ No hardware processing limitations
✓ No additional IT overhead costs
✓ Improved execution time
✓ Meets the needs for more resources to put into production
✓ Ability to run intense algorithms
✓ Good fit with provisioning model
✓ No geographical connectivity issues
✓ High availability of platform
✓ No long waits in queues for processing

of European member states for terms of service

*Below, entry barriers collected from the sustainability questionnaire
✓ The movement of data to from the Cloud
✓ Provisioning time
✓ High learning curve of the programming model and of the execution environment
✓ Infrastructure, and software compatibility
✓ Need to study the APIs and the protocols to use for adapting to the programming requirements of the specific Cloud infrastructure e.g. adapting to the concept realized by the GW module, especially the fact that it does not allow multiple runs on a single Cloud node (multi-threading, multitasking)
✓ Provision of resources is still very low-level and requires cluster administration skills
✓ Choosing the best programming model that fits the Cloud infrastructure (e.g. data flow versus work flow)
✓ Working with non-relational databases
✓ Dealing with systems that are less stable and are not transactional
✓ Setting up a local database to store the inputs of our application, then launching jobs remotely
✓ To learn the development environment and the differences in architecture and development of the Cloud approach with regard to the traditional approach.
✓ Customization of Cloud platform
✓ Finding technical support - crucial for those that do not have an in-house developer or programming experience
✓ The elimination of dependencies from third parties vendors (e.g. ESRI’s ArcGIS scheme) - this demanded re-programming of all the calculation algorithms

<table>
<thead>
<tr>
<th>Opportunities</th>
<th>Threats</th>
</tr>
</thead>
<tbody>
<tr>
<td>✓ Speed-up the research, hence innovation</td>
<td>✓ Vendor (data and applications) lock-in</td>
</tr>
<tr>
<td>✓ Further lowering adoption barriers for SMEs (reduce up-front investment)</td>
<td>✓ Data audit-ability</td>
</tr>
<tr>
<td>✓ Enhancing experiment reproducibility</td>
<td>✓ Jurisdictional issues</td>
</tr>
</tbody>
</table>

Table 2 - Scientific Cloud SWOT Analysis
3.3 Compute Models and Scientific Behaviour

The scientific community has needs that in general are fairly different from typical enterprise customers. Analysing key application characteristics are necessary to understand the feasibility of clouds for scientific applications. As starting point VENUS-C leveraged on the results of the Magellan Report, a project in which Paul Messina, one of the members of the VENUS-C EIAC was fully involved.

The Magellan Report on Cloud Computing for Science classifies scientific workloads in three broad categories based their resource requirements:

- **“Large-scale tightly coupled computations**: These are complex scientific codes generally running at large-scale supercomputing centres through batch queue systems, where Users wait in a managed queue to access the resources. A single job can use up to millions of core hours depending on the scale. These types of applications are expected to take a performance hit when working in Cloud environments.

- **Mid-range tightly coupled**: These applications run at a smaller scale than the large-scale jobs described above. Commonly, Users rely on small compute clusters that are managed by the scientific groups themselves to satisfy these needs.

- **High throughput computing**: Some scientific computing (also large-scale science problems) is performed on the desktop or local clusters and have asynchronous, independent computations. The requirements of such applications are similar to those of the Internet applications that currently dominate the Cloud Computing space, but with far greater data storage and throughput requirements. This community is often concerned with robustness and reliability of jobs over a long-time scale. That is, being able to create a reliable system from unreliable components”.

This structure is very useful as a common starting point for classifying the 27 use cases explored by VENUS-C. Interestingly almost all the cases fit with the first two, while just two cases (Architrave and BLAST) are concerned with some MPI calculation. This reflects the fact that the focus of VENUS-C was, since the very beginning, on the long-tail of science.

The authors of the Magellan Report on Cloud Computing for Science state that “applications in the class belonging to large-scale tightly coupled computations are expected to take a performance hit when working in virtualized Cloud environments. The mid-range tightly-coupled class are good candidates for Cloud Computing even though they might incur some performance hit. Finally, the high throughput computing class may also benefit from the MapReduce programming model by simplifying the programming and execution of this class of applications, in addition to the obvious benefits of Cloud Computing detailed throughout in this document”.

---

In VENUS-C as a complementary exercise another programming model has been explored based on the adoption of COMPSs and GW to explore lightly coupled parallelism. Interestingly the two components – that differentiate in the granularity of job management, proved both to be suitable for the type of applications considered by the 27 cases (see D5.8 Report on Support Scenario).

In addition VENUS-C considered also usage behaviour, not just application or problem type. See Actual Usage Patterns paragraph.

Local Cluster Systems versus Cloud Computing
Local cluster systems and Cloud Computing are best compared when taking into account average utilization rates and expectations on process runtimes. The on-demand model of the Cloud makes it most attractive when compared to a local cluster that is under-utilised since paying for idle cycles is avoided. As per the pros associated with the use of local clusters, one can consider (as per the individual application requirement), a hybrid Cloud could perhaps meet the exact requirements for particular analysis types, for example by providing high memory machines or fast networking, or perhaps legal or data storage issues. In any case, the Cloud market will continue to improve and create new tools and services tailored to the scientific community, as the interest exist, both from the supply and demand end. VENUS-C for example, received a great outcome of 60 responses for the open call (offering very little seed-funding), of which the 15 chosen have all confirmed that Clouds are useful for their community.

In this respect local clusters take a back seat to Cloud Computing, as there is a considerable challenge to consolidating the resources to increase economies of scale and lower operational costs, achieve a high utilization rate and deliver results in a reasonable amount of time, as Cloud Computing has proven to do.

Our User Community
The disciplines represented in the VENUS-C User community fall mainly under the high throughput computing class described above and follow different resource usage models; each discipline typically doesn’t have a fixed frequency of runs over a period of time or the number of runs is not constant throughout a period of time. The VENUS-C User community combines computing needs in high-throughput and high-performance paradigms (including MPI), workflow management, intensive data, and integration with external sources and different types of Users. Most of the scenarios use some degree of parallel computation approach and require accessing external sources of data and network connections. Their execution granularity is highly variable and the applications are data intensive. The below figure provides a summary of the User requirements, their mapping to their respective VENUS-C component, along with their score. The score corresponds to the interest on providing support for the requirement, therefore the higher the score, the higher the interest. The requirements are presented in the two VENUS-C deliverables D4.1 User Community Requirements - first release and D4.2 - User Community Requirements - final release, which offer more details. The table below shows that the
requirements were fairly different, and that the remaining community Users (pilots and experiments) had their requirements met in VENUS-C.

Table 1 – Summary of User requirements scoring and mapping to VENUS-C components

To better understand the application characteristics of our User communities we considered their execution models, profiling of runs (frequency, time, number of CPU and CPU hours per run), memory and bandwidth requirements, data characteristics of the application, persistent disk requirements, characteristics that impact image creation, performance impact of Cloud environment. The sustainability questionnaire asked questions on the following aspects: rough orders of magnitude on running time of a single application run, number of runs typically performs and how frequently, features of Cloud Computing that are of interest, whether they felt the Cloud was technically sound for their scientific community and application, whether the VENUS-C Cloud service satisfies their needs. Below is an analysis of the findings.

Findings

When asked about the run time, the number of runs, and the frequency, our respondents (see below table) provided us essentially with overview of what types of scientific applications fit well with Cloud. The ranges of run time go from 1 minute to forever. The actual run time of their applications decreased when they began using the VENUS-C platform due to increased availability of computational power. The number of runs and frequency are extremely varied; they go from 1 per month to 250,000 a month.
The overall frequency (of a single User) varied quite a bit. For example: one respondent explained that in a typical month they execute 10-100 runs plus 100-40000 runs a couple of times per year. This results in an “irregular” use of the Cloud, rather than a “non-flat” usage model.

Many others seem to follow a “non-flat” usage curve. Another respondent explained that from October – May they perform 0 “zero” runs. While from June-September they execute 3 types of Use cases, first: 1 run per day, second: 24 runs per day and third: approx. 10 runs per day.

To conclude, our respondents described that their usage behaviour in VENUS-C is either quite “spikey” or not very “flat” over a period of time, best viewed depending on the frequency at a month to a years’ time.
How many runs do you typically perform? And how frequently?

<table>
<thead>
<tr>
<th>Frequency</th>
<th>Number of Runs</th>
</tr>
</thead>
<tbody>
<tr>
<td>2/day</td>
<td>2 (every day)</td>
</tr>
<tr>
<td>20/day</td>
<td>20 (every day)</td>
</tr>
<tr>
<td>1-10 / min (200,000-250,000/month)</td>
<td>10-100 runs per 3-months (330/month)</td>
</tr>
<tr>
<td>1000 runs per 3-months (330/month)</td>
<td>up to millions, depending on the library to scan (10,000-100,000/month)</td>
</tr>
<tr>
<td>more than 10</td>
<td>more than 10</td>
</tr>
<tr>
<td>1/week</td>
<td>1/week</td>
</tr>
<tr>
<td>10/month</td>
<td>10/month</td>
</tr>
<tr>
<td>800 runs every few days</td>
<td>800 runs every few days</td>
</tr>
<tr>
<td>Depends on research objectives; however, currently ~50/month.</td>
<td>Depends on research objectives; however, currently ~50/month.</td>
</tr>
<tr>
<td>&gt;500/day</td>
<td>&gt;500/day</td>
</tr>
<tr>
<td>50 every two weeks</td>
<td>50 every two weeks</td>
</tr>
<tr>
<td>10k</td>
<td>10k</td>
</tr>
<tr>
<td>100, 1-3 times</td>
<td>100, 1-3 times</td>
</tr>
<tr>
<td>10-100 runs monthly, 100-40000 runs couple of times per year.</td>
<td>Due to the computing time and the resources available, we will be flexible in this point.</td>
</tr>
<tr>
<td>October – May: 0 runs</td>
<td>October – May: 0 runs</td>
</tr>
<tr>
<td>June–September: Use case 1: 1 run per day (each run is a batch of 112 jobs)</td>
<td>June–September: Use case 1: 1 run per day (each run is a batch of 112 jobs)</td>
</tr>
<tr>
<td>Use case 2: 24 runs per day (single jobs)</td>
<td>Use case 2: 24 runs per day (single jobs)</td>
</tr>
<tr>
<td>Use case 3: Number is not stable. 10 runs per day is an average approximation</td>
<td>Use case 3: Number is not stable. 10 runs per day is an average approximation</td>
</tr>
<tr>
<td>1 unless an exception occurs and the whole job submission process has to be re-started</td>
<td>1 unless an exception occurs and the whole job submission process has to be re-started</td>
</tr>
</tbody>
</table>

Table 2 – Questionnaire answers to run time and frequency
Furthermore, we asked “How do you manage the data for your application? Please check all that apply”. The majority (45-61%) of the respondents selected that their output data needs to be pushed out to a different location, use and expect local disk, large amounts of static data is required on site where the application runs, and input data arrives from offsite.

Figure 4 – Questionnaire Answer – How do you manage the data for your application?

Not all benefits and weaknesses have the same weight when considering where to run an application. For example, our questionnaire asked: “Virtual machines are a popular mode of operation in Cloud Computing that tends to affect some applications in terms of performance. Does this affect your application?” Approximately 40% of the respondents said that virtual machines do NOT affect their application, while of the remaining 60%, approximately 70% said YES, but the other Cloud features are more attractive. Therefore although a number of Users anticipated the performance of their applications would be adversely impacted by virtualization, many felt that the other advantages of Cloud Computing still made it an attractive platform for their discipline.
When asked: “Which 5 of the following Cloud Computing features do you find most attractive?”
The top 5 most selected in order of highest percentage are: Availability to scale up/down depending on the workload at 89%, Access to additional resources at 61%, Short time and cost for setting up new services at 56%, Access to on-demand (commercial) paid resources closer to deadlines tied with cost associativity at 44%, Ability to use software environments specific to my application at 39%.
Figure 6 – Questionnaire Answer – Top 5 Cloud Computing features which are most attractive

- Access to additional resources
- Access to on-demand (commercial) paid resources cheaper than deadlines
- Availability to scale up/down depending on the workload
- Short time and cost for setting up new services
- Ability to scale software environments specific to my application
- Exclusive access to the computing resources/ability to schedule independently of other groups/users
- Easier to acquire/operate than traditional clusters
- Cost effectiveness
- MapReduce/Programming Models/Cloud/Cloud File System
- User interface/Science Gateways
When asked: “*Does the current Cloud service (Azure and/or ENG/KTH/BSC) satisfy your needs?*” Just over 65% of the respondents said YES. The respondents that said NO cited the reasons being that they hadn’t used the infrastructures extensively yet to be able to give an evaluation. Since the respondents’ submission of the questionnaire, other important work has been done on the platform allowing the pilots to use the infrastructure more. Their feedback has been positive.

When asked: “*Do you feel the Cloud is technically sound for your scientific community and the application you use?*” Approximately 85% of the respondents said YES, the remaining said STILL NOT. Those that said STILL NOT noted the reasons: need for more tools for security and tracing of the data, and API to control scaling.

![Figure 7 – Questionnaire Answer – Whether the Cloud is technically sound for each scientific community and the application in use](image)

**Resource Usage Models**

Through the EGI accounting portal\(^{20}\) we can see varied resources usage behaviour of different scientific disciplines over a period of time. The VENUS-C User community also seems to follow these non-flat trends of a period of time. Below is one example of the cumulative normalised CPU time of the Life Sciences discipline from June 2011 to May 2012. All other selected disciplines visualized followed irregular patterns. See Annex B for other graphs.

Usage Behaviour and Cost Effectiveness

The Magellan Report on Cloud Computing for Science describes in what cases moving to the Cloud or beginning with the Cloud are cost effective, below are 2 of the cases:

- **Unknown demand**: In the case of a new project or a new application area where the potential demand is still poorly understood, it can be cost effective to use a commercial Cloud while the demand is quantified and understood. Similarly, if a new project’s prospects are highly uncertain, clouds can be a cost effective option while the long-term fate is determined. In both cases, savings are achieved by avoiding investments in unneeded capacity. For example, our User community includes start-ups - Molplex and Collaboratorio and SME DFRC whom reap perhaps the biggest benefit of Cloud Computing being cost and accessibility.

- **Sporadic demand**: when the demand is highly variable, especially if there are also time sensitive requirements for the service. For cases where the demand must be quickly met, the ability to quickly add additional resources can mean the difference in being able to complete the project or not. Examples are ones with real time-critical requirements.

We can say that the initial investment of acquiring (new) hardware and maintaining the infrastructure with all the overhead costs that come with it is more costly for a User whose computing requirements oscillate throughout the year. This has been often the case with our User community.

For example, one of our scenarios - Civil Protection & Emergencies - Wild Fire would be the extreme case where their runs are not regular throughout the year. From October – May they perform zero
operational runs (maybe they run only for testing or developing new features). Their activity is concentrated in June-September.

The cost-saving could be considered higher/more certain for an application that follows one of more of the following: an irregular pattern (frequency of runs/variable compute demands), has a highly variable compute load, and has an uncertain future. Although, those applications that follow a steadier pattern, rather that the peaks don’t differ too much, certain high-sensitivity factors become evident and drive the economic answer. Aside from their usage behaviour, the Cloud model, according to our respondents is considered a valuable model for the future of scientific activity.

Load Profiles

In a recent analysis comparing Amazon versus self-hosted, the author, Charlie Oppenheimer has analysed the relative costs and sensitivities of both options in the context of different load profiles. By “load profiles,” he means the distribution of demand over the day/month as well as relative needs for bandwidth versus compute resources. He believes the load profile is the key factor influencing the economic choice because it determines what resources are required and how heavily these resources are utilised.

The model he presents provides a simple way to analyse various load profiles and allows one to skew the load between bandwidth-heavy, compute-heavy or any combination. The below image represents his rationale.

![Choosing the Most Cost-Efficient Computing Option](image-url)

**Figure 9 – Choosing the Most Cost-Efficient Computing Option**  
*Source: Charlie Oppenheimer*

His intuitive conclusion is that the spikier the load, the better the economics of the Cloud-on-demand solution is confirmed. The key point he describes is that a comparison of the cost of Cloud hosting versus

---

21 Charlie Oppenheimer, Which is less expensive: Amazon or self-hosted? [http://gigaom.com/2012/02/11/which-is-less-expensive-amazon-or-self-hosted/](http://gigaom.com/2012/02/11/which-is-less-expensive-amazon-or-self-hosted/)  
Feb. 11, 2012
self-hosting needs to be based on the profile of the load. Moreover, it’s not so important where exactly the break-even point is but rather it is most important to know the main sensitivities (e.g. bandwidth cost, CPU load, storage, etc.)

The above analysis would then say that the flatter or less variable the load distribution, the more self-hosting appears to make sense. But, in our case, VENUS-C, a Cloud for scientific applications proves that even those Users who may have a less variable load distribution prefer the Cloud for its many other opportunities and benefits, a decision solely based on cost analysis would be misleading not because it is a very complex area, but mainly because it User requirements are focused on other areas.

According to Hawtin et al., most of the researchers in the UK currently engaged in Cloud Computing are doing so in order to get better performance or new capability. It is therefore possible that increased use of Cloud Computing will lead to more and better science but with an associated increase in cost. With regard to potential advantages, comparisons need to bear in mind the amount of computing that can be brought to bear on the task and hence the time to complete a specific piece of research. Advantages will be most apparent when a large amount of computing power is required for a short time and/or when local facilities lack capacity or availability.

Actual Usage Patterns
Scenarios have been requested to provide real daily usage patterns for their applications during the last period of the project. This information reflects the actual consumption of CPU and relates with the nature of the different scenario usage profiles expected. Figure 11 includes the seven scenario graphs (Y-axis units are CPU hours).

According to this profile, we identify the following real usage patterns:

- **Sporadic peak usage.** Scenarios such as Structural Analysis, Biodiversity and Drug Discovery belong to this pattern. In the case of Structural Analysis, architects make short duration simulations involving a high peak of resources to compute the dynamic load considering different structural conditions and seismic patterns, which require a long post analysis of the results, sometime leading to repeating the simulation under different conditions. Biodiversity offers the processing service for the researchers who want to compute specific predictions of species distributions, with different workload and needs. Drug Discovery may involve several runs to adjust the parameters of the QSAR analysis and then a long analysis time.

- **Oscillatory demand.** Scenarios such as Bioinformatics and System Biology are in this category. In Bioinformatics, long executions are triggered followed by a long analysis process. Then, demand can be satisfied better since the resources are normally provisioned for a long time. System Biology service pre-allocates a set of resources during the experimental times, and are freed once the experiments are concluded.

---

22 Cost Analysis of cloud computing for research - Final report to EPSRC and JISC, R. Hawton, M. Hammond, L. Gillam, G. Curtis, February 2012
- **Plateau of resources.** In some cases, demand is constant, and cloud is used as a way to provide resources without taking the burden of their administration. Scenarios such as Building Information Management pre-allocate resources that are exposed to the community through the portal.

![Structural Analysis Scenario](image1)

![Drug Discovery](image2)

![Biodiversity Scenario](image3)

![Bioinformatics scenario](image4)

![System Biology](image5)

![Fire Risk and propagation](image6)

![Building Information Management](image7)

**Figure 10 – Daily CPU consumption (CPU per hours) of the different scenarios**
“Home is where the Cloud is”

The VENUS-C Scenarios and Pilots have expressed their overall content with their experience with our Cloud and the API’s it offers. VENUS-C has managed to embrace all the diverse User requirements, usage behaviour and offer a great platform where to do great science. The manner in which they use the Cloud does not follow one typical curve, rather they fit into usage curve, some more “spikey” than others, but they’ve all found a place where there are not only requirements met but often exceeded in regards to performance and cost saving aspects. All in all, the scientific behaviour that our User community has is suitable for the Cloud business model, regardless of their usage trends.

3.4 Gap Analysis for Adoption of Cloud Computing by e-Science

Our sustainability questionnaire asked the question: “What do you feel are the major obstacle(s) you have to overcome to be able to use a public Cloud?” The answers act as the gaps our User community encountered. Below is a list of what obstacles the respondents personally feel they need to overcome to be able to use a public Cloud. Notably, financial issues were the most reoccurring obstacle.

- **Legal Issues about licensing and data legal property** (example: data in private and non-private information). Our Italian scenario, CNR (a public body), noted (to the best of their knowledge) that the costs for public Cloud are still not eligible for funding according to the Italian regulation.
- **Financial, personnel (admin) and expertise in moving applications to the Cloud, financial issues, in regards to budget to pay for Cloud services.**
- **For those with Linux interests: flexibility, software compatibility, and portability were noted.**
- **Deployment difficulty** (is still not trivial).
- **Intellectual property rights for research or industry discoveries.**
- **Security when dealing with private data/ data protection.**
- **Availability of a complete and high reliable service, with all the required features.**
- **Technical issues related to porting the applications, rather there is room for improvement to increase User-friendliness.**
- **Clear and complete documentation.**

Finally, below is a great comment addressing an aspect perhaps spoken far too little about:

“There is still a psychological jump needed by the scientific Users to see that Clouds are as reliable as local resources, and in many cases cheaper in cost (this last point is not clearly perceived since many maintenance costs are hidden to the scientist, who does not pay for power or technical staff).” (Ignacio Blanquer - VENUS-C Scenario: Bioinformatics/Structural Analysis)

An Overview

For completeness, we have provided an analysis of the different gaps (technical and non-technical) and challenges facing Cloud Computing today.

Of those that hesitate to consider or move to the Cloud, and those who understand its potential but are clear on what gaps need closing, the following issues arise the most: standards, certification, data protection, interoperability, lock-in and legal certainty.
The European Commission recently published a report called “Advances in Clouds - Research in Future Cloud Computing”\textsuperscript{23}, which details what gaps and open areas are currently present. The table below summarizes the technical and non-technical gaps according to the report.

<table>
<thead>
<tr>
<th>Area</th>
<th>Main Issues</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manageability</td>
<td>Efficiency; interoperability; compensating insufficient resources; boundary criteria</td>
</tr>
<tr>
<td>Data Management</td>
<td>Data size; interoperability; control; distribution; consistency &amp; multi-tenancy</td>
</tr>
<tr>
<td>Privacy &amp; Security</td>
<td>Multi-tenancy, trust, data-encryption, legislation compliance</td>
</tr>
<tr>
<td>Federation &amp; Interoperability</td>
<td>Proprietary structures / de-facto standards; vendor lock-in</td>
</tr>
<tr>
<td>Virtualisation, Elasticity and Adaptability</td>
<td>Elasticity; optimised scheduling; interoperability; resource manageability; rapidly changing workloads.</td>
</tr>
<tr>
<td>APIs, Programming Models &amp; Resource Control</td>
<td>Connectivity; intelligent distribution (code &amp; data); multi-tenancy; enhanced manageability; reliability; ease of use; development and deployment support</td>
</tr>
<tr>
<td>Legislation, Government &amp; Policies</td>
<td>Legislation; governance; licensing; globalisation.</td>
</tr>
<tr>
<td>Economic Concerns</td>
<td>Extended business knowledge; improved QoS management; Green Agenda; energy proportional computing.</td>
</tr>
</tbody>
</table>

Below is a description of some of the gaps mentioned by the questionnaire respondents and also listed in the table above.

**Technical Gap - Data Management** - the actual usage behaviour with respect to file and data access in Cloud systems need to be assessed more carefully, and will therefore help identifying the typical distribution, access, consistency etc. requirements of the individual use cases.

**Technical Gap - Privacy & Security** - Strongly related to the issues concerning legislation and data distribution is the concern of data protection and other potential security issues. New security

governance models & processes are required that cater for the specific issues arising from the Cloud model. In particular sensitive data or protected applications are critical for outsourcing issues. Although, essential security aspects are addressed by most tools, additional issues apply through the specifics of Cloud systems, in particular related to the replication and distribution of data in potentially worldwide resource infrastructures. Even though the data should be protected in a form that addresses legislative issues with respect to data location, it should at the same still be manageable by the system. In addition, the many usages of Cloud systems and the variety of Cloud deployment types imply different security models and requirements by the User. Therefore, traditional authentication models may be insufficient to distinguish between the different actors, in particular in Infrastructure as a Service (IaaS) Cloud systems, where the computational image may host services that are made accessible to Users. In particular in cases of aggregation and resale of Cloud systems, the mix of security mechanisms may not only lead to problems of compatibility, but may also lead to the User distrusting the model due to lack of insight.

**Technical Gap - Federation & Interoperability**

One of the most pressing issues with respect to Cloud Computing is the difference between the individual vendor approaches, and the implicit lack of interoperability. Whilst a distributed data environment based on IaaS is not easily moved to any platform provider (PaaS) and may even cause problems to be used by a specific service (SaaS), it is also almost impossible to move a service / image / environment between providers on the same level. This issue is mostly caused by the proprietary data structures employed by each provider individually. History of web service standardisation has shown that specifications may easily diverge rather than converge if too many parallel standardisation strands are pursued. Moreover, interoperability is normally driven stronger by de facto standards, than by other de jure standardization efforts. In particular Cloud Computing with the strong industrial drivers and the initial uptake already in place has a strong tendency to impel de-facto standards. While innovations between domains usually benefit from an early focus on interoperability, the quest for disruptive innovations within domains benefits from a lower focus on interoperability requirements in this early phase. New policies and approaches may therefore be needed to ensure convergence and thus achieve real interoperability rather than adding to the issue of divergence.

**Technical Gap - APIs, Programming Models & Resource Control**

Cloud virtual machines tend to be built for fixed resource environments, thus allowing horizontal scalability (instance replication) better than vertical scalability (changes in the resource structure) – however, future systems will have to show more flexibility with this respect to adapt better to requirements and capabilities. In addition, more fine grained control over e.g. distribution of data etc. must be granted to the developer in order to address legislation issues, but also to exploit specific code requirements. Cloud systems face issues, for example with respect to description of connectivity requirements, but also to ensure reliability of execution, which is still a major obstacle in distributed systems. At the same time, the model must be simple enough to be employed by average scientific Users. Cloud Computing interfaces and features need to be provided in an easy-to-use fashion for
scientific Users, but should also allow for extended control for more advanced Users by providing new means to manage resources and infrastructure. Development support for new applications has to ensure portability of application (segments) across the network, enabling a more distributed execution and communication model within and between applications. Since Cloud applications are likely to be used by much more tenants and Users than non-Cloud applications, customizability is a must.

Non-Technical Gap - Legislation, Government & Policies
Data, application and services (for example, licensing models) are subject to specific legislation issues that may depend on the location they are currently hosted in, but also applications and services, in particular regarding their licensing models. Legislation issues arise due to the fact that different countries have different laws regarding which kind of data is allowed, but also which data may be hosted where. With the Cloud principally hosting data / code anywhere within the distributed infrastructure, there is a need for new legislative models/means to handle legislative constraints during data distribution. The governance of Clouds needs to be more open to address the actual User requirements better, such as data privacy issues and issues caused by business (process) requirements. Governance solution could also help to select only those vendors providing open-source solutions, thus avoid vendor lock-in. New policies and cross-country regulation are required to deal with jurisdiction, data sovereignty and support for law enforcement agencies.

4. SUSTAINABILITY ANALYSIS
The sustainability analysis describes the current landscape, providing details on strategy models, on our proposed approach, a summary of the business models tied in with our User community. After which, an analysis is provided on different related elements being: technical, organizational, financial policy, legal, and other pertinent elements.

4.1 Strategies for e-Science Clouds
In the last years both virtualisation and Cloud Computing have been steadily gaining visibility in terms of not only expectations but also actual developments, initially from industry but recently also by the research and academic community. Currently, EGI.eu is examining virtualisation and clouds technologies recently publishing a related strategy document, while the two projects StratusLab and VENUS-C have been validating the scientific needs at Infrastructure- and Platform as a Service level. The great interest of research teams in Cloud technology was evident from the highly successful VENUS-C Open Call. However, not just technical feasibility needs to be considered. The Digital Agenda for Europe clearly underlines the need for an EU-wide strategy on Cloud Computing for science that should consider economic, legal and institutional aspects as being done in this study.

In this section, three different strategy models are analysed, namely a:

- Private or community Cloud model for e-Science
- Public Cloud model for e-Science (for example the use of commercial clouds for e-Science)
• Hybrid model, i.e. a mix of the two above-mentioned models, which is also the approach proposed in this report.

In other words we put in practice and analyse a hybrid model of a private/community European Cloud for research (offering both “grid” and Cloud services) combined with external Cloud services offered by one or more commercial providers through a procurement process.

Private/Community Federated Cloud for e-Science

The European computing e-Infrastructure ecosystem has recently started exploring the virtualisation and Cloud Computing offerings. The technological and societal conditions are getting mature enough. However, policy and legal issues such as security, privacy and, trust, and service provisioning including Service Level Agreements (SLAs) also play a crucial role in the wider adoption of the public Cloud services by the scientific communities. The EC, as stated in the Digital Agenda for Europe\(^{24}\), underlines the need for an EU-wide strategy on cloud computing for science that should consider economic, legal and institutional aspects; enablers and barriers for a wider adoption of the Cloud model as one envisioned core component of the future EU digital economy need to be studied and tested in practice. A possible path to overcome many of those challenges is to use a private or community Cloud for e-Science federating national or regional Cloud resources to form a pan-European Cloud infrastructure.

EGI started in 2011 to seriously consider virtualisation and Cloud Computing technologies and performed a series of workshops (entitled User virtualization workshops\(^{25}\)) which were followed during the EGI Technical and Community Forums in September 2011 and March 2012. Other e-Infrastructure DCI projects and especially the European Middleware Initiative (EMI), which is one of the pieces in this distributed computing e-Infrastructure puzzle, contribute with their own strengths and pace towards virtualisation and Cloud Computing. Key issues that still need to be worked out are how National Grid Initiatives (NGIs) and EGI can move towards a Cloud Computing model to support research and data-intensive science and if this service should be provided by federations of NGIs.

There are quite a few NGIs that already offer pilot Cloud services, in most cases starting with simple IaaS services such as access to raw virtual machines. In this way, new, more User-friendly interfaces are being deployed. Successful tests from EGI and StratusLab have also been made with the grid infrastructure being silently extended with underlying Cloud resources. More resource providers will be evaluating the virtualization and Cloud-related technologies and services as the DCI e-Infrastructure projects produce more concrete results. A mix of grid and Cloud interfaces and services is thus expected to be available in the near future so as to be able to serve both existing and new Users, experienced and inexperienced ones.


In March 2012, EGI published its draft EGI strategy for public comments\(^{26}\). The main points from this document relevant to this discussion area are the following:

- To develop and promote technologies for federating the emerging Cloud resources.
- To develop and promote technologies to merge existing grid services with virtualization and Cloud resources.
- To inject the flexible on-demand Cloud technologies and services tuned to the specific needs and workloads of the scientific community.

A private/community Cloud for e-Science can thus eliminate the privacy, trust, legal and other challenges that come with public Clouds. This can be either transformed by federating national Cloud resources and/or by also having some centralized resources (the latter though seems not to be considered by EGI). In this way the current EGI grid will be gradually transformed into a private/community Cloud offering both Grid and Virtualisation/Cloud services. The disadvantages of this solution would be the lack of commercial (public cloud) resources that can easily extend the infrastructure in case of severe stress and full utilisation of the infrastructure and the fact that the so-called long-tail of scientists that cannot afford some of the mechanisms of the grid, which may be considered complex, is not easily integrated into such a private/community Cloud.

However, the VO-based resource allocation scheme of the grid and the distributed control of the over-utilised resources (owned by the NGIs or communities) is certainly a weakness (also identified in the EGI strategy document\(^{27}\): “No clear top-down resource allocation model or process comparable to commercial cloud services due to the bottom-up ownership and allocation of resources directly to specific local research communities”

There are of course some general-purpose VOs (catch-all VOs) for the individual researchers of the long-tail of science, but the grid model is such that EGI pushes such individual researchers to organise themselves with other researchers in the same application area around the world, build their own VOs and bring their own resources to support these new VOs. But certainly for many cases, this does not work well. So, again this advocates for a central pool of resources, possibly cloud-based resources from commercial providers (or private), and have a hybrid infrastructure of virtualised federated grid resources enhanced by centrally-controlled public cloud resources. Some EGI efforts on such on-demand/centralised control type of jobs -the so-called “short-lived jobs” (jobs that had to be executed immediately otherwise failed) have already faded away.

The use of public Cloud for e-Science has already been analysed in chapter 2 and especially section 2.2. This approach involved the direct usage of public Clouds (for example commercial providers) such as Amazon or Windows Azure for scientific usage. This type of model has already been successfully tested in the VENUS-C project by a wide variety of applications and with very good results, as also shown by the


\(^{27}\) The European Grid Infrastructure (EGI) strategy paper for Horizon 2020 ‘Seeking new horizons: EGI’s role for 2020’, http://go.egi.eu/EGI2020
sustainability questionnaire. Such a model thus allows the long-tail of science to use the Cloud and offers all the benefits of commercial Clouds (see also the SWOT analysis).

Still, there are several issues that need to be taken into account as already analysed, namely vendor lock-in and interoperability when moving to a new Cloud, privacy, trust, jurisdiction, SLAs and quite a few others. Such issues need to systematically studied and resolved and might act as an impediment for their wider adoption.

The European Space Agency (ESA) and CERN kicked off an initiative on Scientific Cloud in Europe\(^\text{28}\), which was later named “Helix Nebula - the Science Cloud”\(^\text{29}\) and also joined by the European Molecular Biology Laboratory (EMBL), which seems to focus on the fully-outsourced model of the computing resources to public Clouds. Even if this can be a reasonable approach from the economic and technological point of view, some of the security and legal risks may delay a wider adoption by the whole EU community. Furthermore, it is obvious that Helix Nebula deals mainly with the large and powerful Users, directly opposite from the long tail of science that VENUS-C had been targeting. StratusLab with some of its partners look on industrial uptake of their software efforts as part of Helix Nebula.

Sustainability Approaches for e-Science Clouds

One of the most notable approaches for an easy adoption of Cloud Computing is the deployment of a hybrid Cloud, where private resources and public ones coexist in a federated environment. The main reason is due to the existence of previous investments – computing and storage resources as well as deep know-how of disciplines needs – and the potential hidden costs or security and legal issues in a pure public Cloud. The challenge here is about understanding how to implement a hybrid model mixing a federated private/community Cloud for research with external public Cloud services offered by one or more commercial providers, either adding resources to a grid or/and offering direct “Cloud” services.

If we consider the current scientific landscape a federated community Cloud could be an evolution of the pan-European e-Infrastructure (such as EGI) or other national organisation (federating resources at local or regional level), gradually implementing all the technologies and services that have been tested in the forerunner Cloud projects.

This private community research Cloud will act as a first level of aggregation of resources. However, as pointed out above, this internal cloud is usually not convenient for individual scientists or small teams of researchers – the long tail of science in the long run (due to the legacy and inherent operation of the grid requiring that such individual users team up and gradually bring their own resources into the infrastructure). Furthermore, during major disciplinary service challenges or other unexpected events or crises it won’t be able to guarantee a sustained level of service for all communities, and therefore peak traffic demands (the so-called cloudbursting) or in cases of unavailability of resources can be served on-


demand by commercial providers such as Amazon or Microsoft but also by other EU-based initiatives that may become reality in the near future. This implies either transparently extending the grid infrastructure with underlying leased Cloud resources or extending the community Cloud by pooling more on-demand Cloud resources to it. Of course there are a series of technical and non-technical issues to resolve about platform compatibility, open interfaces, interoperability and standards, data replication, legal and financial issues, and more. A competitive procedure (procurement) for the selection of one or more commercial providers should be also added to the list of challenges, together with the associated administrative issues, an area where GÉANT and the DANTE/NRENs’ experiences in procuring telecommunication services will be valuable.

A well-balanced hybrid model could offer the advantages of both worlds, i.e. the research e-Infrastructure providers and the commercial ones; that is the sophisticated, secure environment and all the vested know-how of the first with the elasticity, flexibility, User-friendliness and cost-effectiveness of the latter.

Other options, equally relevant are related to the direct access of individual scientists or small research teams to the Cloud resources they need without intermediation. This pose a challenge in terms of funding mechanisms: currently the EU FP7 Framework and the funding-by-project-calls mechanism foresee an upfront decision on all possible costs that put the researchers in a more safe position (from an administrative perspective) when they bought bare metal instead of computing services. The next revision of the legal framework (Directive 95/46/EC) and the European Cloud Partnership may renew the interest for such approach.

Summary of considered Business Models
In this sub-section the business models supported by VENUS-C are considered and generalised. The first case, the researcher already has access to a campus, regional, national or European HTC computing e-Infrastructure (such as an NGI or EGI). The latter e-infrastructure entity or entities decide to procure services from public Cloud providers to extend their infrastructure (e.g. for cloud bursting) either on-demand or steadily for a period of time possibly for some specific Users (long-tail of researchers/scientists). This may mean that either the public Cloud resources are transparently integrated into the in-house campus/national/EU e-Infrastructure (in which case the User does not notice any change) or the User is asked to access the public Cloud possibly through a new authentication mechanism and a new set of tools/interfaces. In all cases, the flow of money is through the research e-Infrastructure entity and the public Cloud providers. The two subcases are both part of a hybrid model involving a private or community cloud and a public cloud, although the first is transparent to the User. A VENUS-C platform can also be part of the research e-Infrastructure to facilitate the hybrid model and interlink with the public provider.
The second case is similar to the first besides the fact that a 3rd party broker is used to access the public Cloud providers. In this case the flow of money takes place both between the broker and the Cloud providers and also between research e-Infrastructure and broker. VENUS-C components can again play a role either as part of the broker or as before.
In the third case, the researcher, who is usually a member of an academic or research institution, interacts directly with one (or more) public Cloud providers, either using credits (that have already been agreed with the Cloud providers) or through an institution credit card (e.g. pre-filled). The money flow can take place either between the researcher and the Cloud provider or between the research e-Infrastructure and the Cloud provider. A VENUS-C local instance (or components) can also be installed in the laptop or computer that is used to access the public Cloud. Individual access of a researcher to a Cloud provider (with his own funding) is of course possible but not directly in the scope of our study.

![Figure 13 – Individual scientist with direct access to public cloud model](image)

4.2 Sample of VENUS-C Business Models
A sample of business models is provided below with the aim of demonstrating current examples of new services that have been created around the scenarios at a commercial or pre-commercial level.

**Bioinformatics**: The availability of a service for speeding-up the processing of BLAST/BLAT and similar processing tools is of key importance for companies working in area of bioinformatics consultation. Many SMEs do have their own special tools that they use either in premises or offer their customers. In many cases, companies offer two-levels of services: 1) free access with reduced functionality and support on one side, and 2) subscriber access (with payment) to an enhanced tool and improved support. In most cases, final Users (customers) are normally researchers. Main weakness of such companies is on the availability of resources. An example is B2GO, software exploited by BioBAM[^30], a newly formed bioinformatics company that has emerged from an academic environment that needs computing resources for BLAST and BLAT. The **subscriber model** can dedicate part of the funding for paying the access to resources.

[^30]: http://www.biobam.com/
As the developer of the bioinformatics service in VENUS-C, the Universidad Politécnica de Valencia can exploit outputs through the release of a supported version of the service, and through the monitoring and maintenance of the service on a public cloud. The university is already setting up an agreement with BioBAM to pilot a business model experiment during the additional year of VENUS-C (June 2012-May 2013). The university will use mainly the Generic Worker for accessing the public infrastructures (possibly COMPSs + OVF4ONE for other IaaS public infrastructures), CDMI service for standard access and accounting service to differentiate individual Users in each deployment. Bioinformatics will mainly focus on the public infrastructure model.

**Structural analysis:** The Universidad Politécnica de Valencia outputs bring value-add in terms of being able to provide enhanced services running in the cloud with the availability of intensive resources. The university has investigated privacy issues for exploiting this model. It must be noted that a broad test of the different configuration for dynamic analysis, which takes time, cannot be exploited easily by including it in desktop versions, since it could create customer rejection due to long execution times and the need for updating the regulations related to the experiments. The university is therefore focusing on providing this kind of high-detail analysis services only on-line. The hybrid model is also very interesting in this case, since a basic level of resources can be used to support training and higher-education activities. Specifically, the university will leverage open-source components developed to offer such a service (Generic Worker for accessing the public production infrastructure - Azure, COMPSs + OVF4ONE for training and education private infrastructure – setting up tailored resources that draw on the experiences of BSC, KTH and Engineering, CDMI proxy for accessing data from a single client and accounting service to track usage). The applications adapted and the processing services are directly exploitable for the points described.

**3D Rendering Visualization:** Immersive rendering functionality is extremely important when it comes to promoting a design project. Architects and investors promote their project through the construction of complex rendered images and videos which require high computational power. Many architectural firms outsource this task because they do not have an internal system that allows them to quickly create a realistic view of the project. In our particular pilot case, immersive rendering provides a complete view of architectural concepts, enriches the CAD project published online, and helps the jury to define the best project in a call for bids system. There are quite a high number of applications for 3D modelling and architectural rendering. Collaboratorio selected the candidate software based on the outcomes of a survey completed by its Green Prefab community. The expected advantages are: speeding up the time needed to run the entire simulation; no additional IT provisioning and maintenance costs; pay for use in a transparent environment; scalability of the service (large number of computational resources available on demand); and ability to solve problems with more accuracy. Green Prefab is a start-up spin-out of VENUS-C that is providing the new cloud-based service, as described below.
The three new complementary services for architects and engineers offered through HUB-ENGINEERING, (HUB-E) are hosted on a dedicated Green Prefab portal (www.hub-e.com), where the three services with related ‘communities’ and chosen business models are advertised:

- 3D Rendering visualization – Luxrender, free to all Users
- Eco-efficiency analysis – EnergyPlus, Royal Danish Academy (VENUS-C pilot) with pricing on project
- Structural analysis – Architave, Universidad Politécnica de Valencia with pricing on project.

Each dedicated section on the ‘Services’ page, illustrated below, shows the compliant software.

**Figure 14 – Luxrender**

**3D rendering visualization**

**Figure 15 – EnergyPlus**

**Eco-efficiency analysis**
4.3 Technical Elements
As already analysed in section 3.3, the Cloud may not be suitable for all types of applications. “High throughput computing” type of applications are more dynamic and workloads typically consist of large number of independent computing tasks that can run in parallel. Sometimes they are referred to as “loosely coupled” computation tasks, compared to the “tightly coupled” HPC ones that require high-communication between each other and thus fast interconnects between the computing nodes. HTC workloads cannot be handled by one HPC facility and thus require sharing of resources (including virtualised ones) that are distributed. In short, the size of workload, nature of jobs and resource sharing are the distinguishing factors and Cloud Computing is best suited for such (HTC) tasks. Characteristic examples of HTC requirements can be found in Table 3 (section 2.2) such as Map Reduce, Data Flows, Data-driven Multiple Job Execution, etc.

Nowadays, enhanced or “high-class” Cloud or other HTC infrastructures are also capable of handling some portion of the “tightly-coupled” computations tasks (in the Magellan report referred to as “mid-range tightly coupled” ones – see 3.3).

A VENUS-C internal document on HPC in the Cloud analysed the VENUS-C Platform components, especially the programming models, in order to describe how certain types of applications typically hosted in supercomputing centres and with no large scale requirements in terms of interconnectivity and I/O, could be effectively ported to the VENUS-C Platform. The findings of the report determined that the VENUS-C cloud platform can provide the mechanisms and tools for high-throughput and data-intensive applications that, while not typical of the traditional model of supercomputing centres, have increasing needs for computation and storage resources.

Furthermore some baseline of interoperability or some standard interfaces will help avoiding the vendor lock-in and promote sustainability of the selected approach. In order for an application to be sustainable in a virtualized Cloud environment, it has to satisfy the requirements of the specific application and
related User community. Having Users “technically satisfied” is probably the first thing that they will look, after of course securing their funding which is an absolute prerequisite.

4.4 Organisational Elements

Having scientific Users running their applications on public Clouds requires a series of organisational changes as the traditional way of doing their research has been to get their computation through self-hosted computers or campus-regional and/or national-European research e-Infrastructure providers. Major organisational changes required have to do first of all with the funding clauses, obligations and cycle. In most cases funding for computation is given to campus or national computing e-Infrastructure providers such as University and Research Centres computing centres or National Grid Initiatives. Funding agencies thus need to allow paying for public Cloud services either centrally or possibly in a distributed way. The campus, regional or NGI entity can in this case act as a demand aggregator for corresponding Cloud services (see also related business models). In case the funding agencies also allow individual researchers to use such services (though some pre-paid credit cards or other way e.g. PayPal), the whole funding cycle has to change and some way of measuring the needs and ethical usage of the services needs to be installed. The experience from the pilots (and the related questionnaire) reveals that in some organisations or countries using public Cloud Computing services is currently not eligible or that researchers are not able to make the necessary credit card reimbursement claims.

Depending on the strategy, funding option, and business model used for paying Cloud services, corresponding governance models need to be in place. The governance models need to enforce committees for reviewing, accepting, monitoring the needs and requests of the different entities in cycles (e.g. annually). These can be again at a campus, regional, national level, or even possibly at European level, in case there is an entity such as EGI.eu that undertakes to aggregate the demand for the whole European HTC research communities (possibly per application or Virtual Organisation) and order public Cloud services. If a hybrid strategy/model is agreed, for example a mixed community Cloud with public Cloud services for cloudbursting, it would be much easier for the European entity (such as EGI.eu) to implement this scenario (than for all HTC VOs in Europe).

The governance models need also to provide for legal and financial experts (besides technology/technical experts) to better understand the offerings and SLAs from the commercial entities and select the optimal offer(s). Possibly new profiles of experts need to be developed and promoted for this new era. The hybrid model will provide for this gradual transition, starting the just the cloudbursting needs outsourced to commercial Clouds, better understand the financial, legal and policy implications along with related human resources expertise, and get better organised overall for this change.

Sustainability of future approaches has once again to satisfy the needs of the Users, but also enable the funding agencies to control and guarantee the efficiency of their subsidies. A gradual transition to a hybrid model possibly followed up by public clouds approach will certainly give some more time of testing and better understanding the risks while actually using public Cloud services.
4.5 Financial Elements

The most important financial element that needs to be understood is the cost aspect, i.e. being able to compare the costs of public clouds with the costs of the research e-Infrastructure providers. However, as identified in the e-FISCAL project\(^{31}\) such a cost comparison is not trivial. First of all, most of the Cloud vendors offer virtual machines while currently e-Infrastructure providers (besides VENUS-C) offer computing cores. The e-Infrastructure providers such as EGI are also very distributed with a wide variety of computing facilities (ranging from normal servers to high-end clusters with fast interconnects and MPI support). So comparisons need to be based on some kind of average performance based on some representative selection of benchmarks that can at least satisfy a variety of basic User requirements. Even when this is accomplished, assessing the costs of the e-Infrastructure providers might be a mission impossible, given again the distributed nature of the infrastructure, such as the high variation of machines, the different centres or NGIs that run it and the lack of financial data in those centres. And of course one can compare the costs (from the research e-Infrastructures) with prices (from Cloud vendors) that might include a profit margin or even a loss (loss leader pricing strategy selling below cost\(^{32}\)).

Aspects such as availability of the services, utilisation and other SLA obligations are also important and need to be taken into account. In fact, utilisation is a crucial factor for determining which infrastructure is more cost-effective. Further, the large vendors are constantly innovating, and charging new models are occasionally introduced that significantly affect the economics and thus also make comparisons outdated in the medium term\(^{33}\).

Reviewing the state-of-the-art\(^{34}\), reveals that such a comparison is indeed not straightforward. There are a few cases comparing outsourced services (e.g. to Amazon) with self-hosted e-Infrastructures, the vast majority being from the US (such as the Magellan report or US universities). Although vendors change their prices quite frequently, it is indicated that research e-Infrastructures aggregating demand and having high utilisations are most cost-effective in the long run (especially with high depreciation rates e.g. 4-5 years). On the other hand, for spiky traffic for only a fraction of the year or with low utilisation rates this might be opposite. Such an example is the Wild Fire application scenario from VENUS-C running for less than 5 months a year. So once again, such a comparison is not trivial and requires a careful study from experts with different expertise such as accountants and technical experts for benchmarking.

\(^{32}\) http://en.wikipedia.org/wiki/Loss_leader
\(^{34}\) www.efiscal.eu/state-of-the-art
Besides the cost comparison issue, other financial elements including taxing and invoicing issues, applicability or eligibility of such taxes to be paid, especially if this is abroad or even beyond Europe (e.g. US).

So the financial aspects pose quite a few challenges for the sustainability of the Cloud model and more work for experts in order to make the proper analyses and studies to come up with safe results. A network of financial experts needs to be built, promoting the clear documentation of the costs, possibly with proposed cost models and rules of thumb need to be prepared (for costs per core, CAPEX/OPEX ratios, etc.) R. Hawtin et al. recommendation is that funding bodies/grant reviewers at national level work together with relevant stakeholders such as special interest groups and national research and education networks (NRENs) to benchmark the cost of computing at institutional level, while also developing and publishing a knowledge base of the use of cloud computing for research and investigating further use of reserved instances.35

4.6 Policy Elements
As identified in VENUS-C deliverable D3.9 a Cloud Strategy is being referenced in the Digital Agenda for Europe communication and for this to happen all the related issues and elements need to be well understood. Cloud Computing is certainly the future and its gradual testing and validation will trigger clear results and actions for all related stakeholders, for example policy ones including funding agencies, e-Infrastructure providers and Users. Horizon2020 is a unique opportunity for Europe to test the Hybrid model, promote Public Private Partnerships and validate public vendors. This will also safeguard the investments of national and European funding agencies in the HTC e-Infrastructures, possibly having a depreciation of 3-5 years, but on the other hand also promote innovation and make current e-Infrastructure services more User-friendly, elastic and attractive.

Sustainability is thus depending on the selected model and the involvement of all related stakeholders in these studies, strategies and analyses, once again satisfying the needs of their end-Users.

4.7 Legal Elements
A legal analysis of scientific Clouds has been performed in VENUS-C deliverable D3.7, including also contribution from an external lawyer, namely Paolo Balboni who is an expert in the area. The basic points of this analysis were the following:

- Data Protection Law On The Cloud: Personal and Sensitive Data
- The role of data controller and data processor
- Risk of changes in jurisdiction
- Data Security
- Privacy & Confidentiality Laws
- Identity & Access Management
- Intellectual Property Rights

35 Cost Analysis of Cloud computing for research, op cit.
• Service Level Agreements
• European Valued Added Tax regulation

The analysis included special references to Scientific Clouds and details can be found in the VENUS-C deliverable D3.9 – Potential Legal Issues.

4.8 Other Elements
Other elements that affect the sustainability of hybrid or public Cloud services are the ones of training and vendor lock-in that have been already identified. Cloud Computing paradigms and programming models required quite deep technical expertise and this has to be gradually acquired. Training and related human profiles with appropriate expertise need to be trained. For example, the current systems administrators of e-Infrastructure providers may need to be transformed to Cloud experts (at the IaaS or PaaS level) understanding the offerings of the vendors and making the orders on behalf of the community or possible move up the stack and acquire expertise in the adaptation of applications in the Cloud. The experience from the pilots (and related questionnaire) confirms also this point on the required training for Clouds36. The SIENA initiative and its roadmap will help capitalise on interoperability standards implementation and pave the way for future directions, including those pursued by VENUS-C, as a reference for all relevant stakeholders. Standardisation or at least interoperation between Cloud providers is also very important as already analysed higher level APIs such as the VENUS-C ones, operating on top of multiple providers and enforcing the hybrid model should be promoted and further developed in the future (especially as part of the Horizon 2020).

36 See also Cost Analysis of Cloud computing for research, op cit., which draws similar conclusions.
5. VENUS-C SUSTAINABILITY AND EXPLOITATION STRATEGY

This chapter is dedicated to the VENUS-C sustainability and exploitation strategy. Here, the exploitable assets are described. The sustainability and exploitation plans of each of the VENUS-C partners are presented. They include a description of the principal VENUS-C components they will exploit, what the VENUS-C selling factors are according to them, and information on their principal customer base and network.

5.1 VENUS-C Sustainability Strategy and Exploitable Assets

During its 24 months of operation, VENUS-C produced significant tangible and intangible assets as described below.

**Tangible Assets**

The tangible assets – listed in below Figure- that have been released with an open-source licence are publicly accessible from the home page of the VENUS-C web channel (www.venus-c.eu, see Table below), linking directly to http://resources.venus-c.eu with interlinks to related documentation provided from the same access point. The open-source components are therefore available to both the current User community, VENUS-C partners as well as to the much broader open-source community. The documentation provided is complemented with the dedicated eTraining environment, which is provided through the VENUS-C web channel, is designed to enable autonomous use through a series of dedicated material, which will undergo enhancements in the near future reflecting feedback received. The value proposition around these assets is also captured and is being used to broadcast them to the wider community to foster uptake.

<table>
<thead>
<tr>
<th>Tangible Assets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generic Worker (PMES and programming framework) – for job management</td>
</tr>
<tr>
<td>COMPSs (PMES and programming framework) – for job management</td>
</tr>
<tr>
<td>CDMI Proxy – for data access</td>
</tr>
<tr>
<td>Accounting and billing service</td>
</tr>
<tr>
<td>Traffic Redundancy Elimination (TRE)</td>
</tr>
<tr>
<td>OVF4ONE -Software component- gateway from OCCI to Open Nebula OCA Interface</td>
</tr>
<tr>
<td>Seven Scenario Applications (Wildfire, RainyCloud, Bio Cloud Tools, Architrave,</td>
</tr>
<tr>
<td>Green Prefab HUB-E, COSBI, QSAR)</td>
</tr>
</tbody>
</table>

Table 4 – VENUS-C Tangible Assets

The below figure is a snapshot of VENUS-C website, where the Open Source components can be downloaded (top right-hand corner)
Figure 17 – VENUS-C Channel snapshot of OS downloadable components

Figure 18 – VENUS-C Channel snapshot of OS downloadable components - Internal page
Intangible Assets

The intangible assets are also publicized on the web channel and summarized on the home page, which focuses on the User communities. These intangible assets range from new knowledge and experience acquired, some of which has fed into the educational training environment, to standards implementation showcased through dedicated events and visibility at multiple levels, which is also documented on the channel. An important achievement is the inclusion of VENUS-C in the Digital Agenda for Europe.

The VENUS-C project has a differentiating value to promote through its great knowledge (know-how) and actual experience lived, collected and documented throughout these 24 months. The knowledge we have gained through our User-centric approach by meeting the vast and varied User community requirements, successfully on-boarding them and their applications to our Cloud platform, and finally the lessons learned are clearly our key long-lasting strength. To highlight, our experience is very rich as our User community includes Users from academia and research structures and SME’s. As a research project, VENUS-C’s deliverables, training material and the sustainability approach described in this very deliverable should be also be considered assets. They represent the real-world experience and a careful study of if and how to set-up a scientific Cloud with a focus on the long tail of scientific Users.

<table>
<thead>
<tr>
<th>Intangible Assets</th>
</tr>
</thead>
<tbody>
<tr>
<td>eTraining environment with forthcoming packages for autonomous end-user support</td>
</tr>
<tr>
<td>eTraining environment with educational services for the wider community</td>
</tr>
<tr>
<td>Project sponsor credibility and reputation in the marketplace</td>
</tr>
<tr>
<td>Commitment to relevant standards, including high-level cloud standards and established protocols, as well as VENUS-C CDMI Client in the EGI Federated Clouds Test Bed</td>
</tr>
<tr>
<td>In-house and peer-review publications on the main assets accrued</td>
</tr>
<tr>
<td>Visibility in the European landscape, including the Digital Agenda for Europe</td>
</tr>
<tr>
<td>General advancement of knowledge (experience and know-how)</td>
</tr>
</tbody>
</table>

Moreover, VENUS-C has made great efforts to show case our work. The inherent visibility is also an important non-physical asset. VENUS-C has a very rich website with many members, we have a good list of publications, papers, and articles, and our “network” is not indifferent. Worth highlighting are the articles written on VENUS-C and recently published on the CORDIS websites and echoed as a feature on the Digital Agenda for Europe. VENUS-C is producing a series of publications on the main technology

---

37 VENUS-C eTraining documentation: http://www.venus-c.eu/Content/eTraining.aspx?id=860fb33b-86e0-45de-9f5d-68b2d7f5f4b
39 Bringing open, user-centric cloud infrastructure to research communities: http://ec.europa.eu/information_society/digital-agenda/research/bringing_open_user.htm and SMEs empowered by a pioneering approach to cloud computing: http://ec.europa.eu/information_society/digital-agenda/research/smes_empowered_by_a_pioneering.htm
advances and the User communities with the aim of capturing the tangible assets and to raise further awareness to the wider scientific and business communities. Recently, our Wildfire scenario and its application has been the centre of attention. Hostingtecnews.com published an article in April on the Wildfire application describing it as “new service has potential for many other regions across the globe affected by devastating fires.”

This is a great achievement for VENUS-C, adding to our list of exploitable assets. In addition, the article has contributed to the positioning VENUS-C in the cloud landscape (PaaS).

Our exploitable assets are not only based on valuable knowledge and great visibility, but also exploitable assets which include specific software components. Our software components carry with them the advantage of open source codes, which creates the special opportunity for future cooperation and continued development.

In regards to what will happen after May 31, 2012, the project will end, but all partners who are involved in providing the platform and services will continue to do so for an additional year on a best effort basis.

The following table provides an overview of the exploitable assets mapped to the VENUS-C partners:

### Exploitable Assets

<table>
<thead>
<tr>
<th>COMMERCIAL EXPLIOTABLE R&amp;D Results (Software Components)</th>
<th>ENG</th>
<th>OGF.eeig</th>
<th>COLB</th>
<th>UNEW</th>
<th>UPVL.C</th>
<th>KTH</th>
<th>COSBI</th>
<th>BSC</th>
<th>CNR</th>
<th>MIC-GRT</th>
<th>MRL+</th>
<th>EMIC</th>
<th>IIT</th>
<th>AEGEAN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generic Worker (PMES and programming framework) – for job management</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>O</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
</tr>
<tr>
<td>COMPSS (PMES and programming framework) – for job management</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>O</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
</tr>
<tr>
<td>CDMI Proxy – for data access</td>
<td>E</td>
<td>O</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
</tr>
<tr>
<td>Accounting and billing service</td>
<td>O</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
</tr>
</tbody>
</table>

**Commercial Expliotaible R&D Results (Scenario Applications)**

<table>
<thead>
<tr>
<th>Wildfire</th>
<th>ENG</th>
<th>OGF.eeig</th>
<th>COLB</th>
<th>UNEW</th>
<th>UPVL.C</th>
<th>KTH</th>
<th>COSBI</th>
<th>BSC</th>
<th>CNR</th>
<th>MIC-GRT</th>
<th>MRL+</th>
<th>EMIC</th>
<th>IIT</th>
<th>AEGEAN</th>
</tr>
</thead>
<tbody>
<tr>
<td>RainyCloud</td>
<td>E</td>
<td>E</td>
<td>O</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
</tr>
</tbody>
</table>

**General Advancement of Knowledge (Experience and Know-How)**

<table>
<thead>
<tr>
<th>OpenNebula for eScience</th>
<th>ENG</th>
<th>OGF.eeig</th>
<th>COLB</th>
<th>UNEW</th>
<th>UPVL.C</th>
<th>KTH</th>
<th>COSBI</th>
<th>BSC</th>
<th>CNR</th>
<th>MIC-GRT</th>
<th>MRL+</th>
<th>EMIC</th>
<th>IIT</th>
<th>AEGEAN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Azure for eScience</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
</tr>
<tr>
<td>Porting of application in to the Cloud</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
</tr>
<tr>
<td>eScience Computing environment</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
</tr>
<tr>
<td>Access to experiences of all VENUS-C user communities (pilots + scenarios + experiments)</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
</tr>
<tr>
<td>Training Material/Documentation</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
</tr>
</tbody>
</table>

**Partners**

<table>
<thead>
<tr>
<th>PARTNERS</th>
</tr>
</thead>
</table>

**Table 6 – Exploitable assets mapped to partners**

### 5.2 Partner and Pilot Final Comments on VENUS-C

The chart below captures VENUS-C through the eyes of the consortium and the platform Users. The chart is a collection of the final comments on the project by the partners, partners with a dual-role of scenario and the pilots. It is a collection of the partners’ comments on exploitation and Users comments on key benefits perceived.
VENUS-C: www.venus-c.eu

<table>
<thead>
<tr>
<th>ENG</th>
<th>VENUS-C represented a major investment in Cloud Computing technology by Engineering. The experience and know how acquired as well as the accounting framework developed will be used as extension of the CLOE services in the ENG production environment.</th>
</tr>
</thead>
<tbody>
<tr>
<td>OGF.eeig</td>
<td>Trust-IT has built a web-based forum, communication tools and training modules integrated into the web channel to support user communities leveraging resources after the end of the project cycle and beyond. These tools also include in-depth analyses of the cloud computing and standards landscape. Through the SIENA initiative, Trust-IT has built strong relations with the standards organisations that have developed standards implemented in VENUS-C (e.g. OGF, SNIA, DMTF) and has actively supporting VENUS-C participation in the cloud plugfest events, thus providing a solid foundation on which to build growing participation in Europe and globally. Trust-IT has spearheaded the Cloudscape flagship workshops as a sustainable high-level policy event that is expected to increase in importance as cloud gains traction, simultaneously analysing market adoption trends and identifying a wide range of stakeholders.</td>
</tr>
<tr>
<td>KTH</td>
<td>KTH has adopted programming models offered by BSC and EMIC and added them to its service portfolio. The work on CDMI-Proxy has resulted in a publication and is being supported beyond the VENUS-C project by KTH as one of its research areas. KTH has now (1) Higher productivity - the self-provisioning nature of the cloud model offered by VENUS-C has allowed to improve on the relative support effort per a client, (2) Increased cost-efficiency - additional load generated by the VENUS-C applications and spread over hardware nodes via virtual machines, increases the efficiency of the hardware usage.</td>
</tr>
<tr>
<td>IIT</td>
<td>VENUS-C offers a unique cloud platform for the benefit of end users. IIT was able to leverage the platform to develop a traffic redundancy elimination algorithm and system. The algorithm is tailored for the cloud environment and enables the acceleration of traffic to the end users while reducing cloud CPU costs usually associated with such tasks.</td>
</tr>
<tr>
<td>BSC</td>
<td>Assets developed in VENUS-C will be leveraged in the EU_Brazil OpenBio project that will leverage on the VENUS-C COMPSs Framework to enhance the porting and execution of the scenarios in the Cloud infrastructure in a seamless way with regard to the underlying infrastructure. The COMPSs programming model will enable the users to program complex workflows without the use of any API leaving to the COMPSs runtime the responsibility of efficiently scheduling the parts of this workflow on the available resources and of interfacing with the data storage in order to retrieve the input data and publish the results.</td>
</tr>
<tr>
<td>MRL, EMIC and MIC-GR</td>
<td>Real Benefits Experimented area: Interaction with a reference set of the long tail of e-Scientists and deployment of their applications on a commercial cloud computing infrastructure. Concrete Result: Identification of requirements for the implementation and deployment of scientific applications on the public cloud resulting in the development of interoperable middleware components offering corresponding services.</td>
</tr>
</tbody>
</table>

PARTNERS AND SCENARIOS (dual role)
<table>
<thead>
<tr>
<th>PILOTS</th>
<th>VENUS-C: Sustainability Perspectives of Scientific Clouds</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>UPVLC</strong></td>
<td>UPVLC leverages VENUS-C for offering Bioinformatics and Structural Analysis scenarios as Software as a Service, increasing business opportunities through new services and outsourced computing that support them.</td>
</tr>
<tr>
<td><strong>COLB</strong></td>
<td>The key benefits we perceived are: (1) Enabling massive communities to cloud computing resources Green building industry is globally asking to hundreds of thousands professionals to engage the status-of-the-art on engineering analysis (especially on energy efficiency). They are now testing their engineering algorithms over articulated sets of 3D drawings with attached metadata. Variables on those 3D models create millions of possible solutions and designers can’t wait so long for outputs. Designers’ workstations are not anymore enough. (2) Increased problem size. In Building Environmental Performance Simulation the ‘problem size’ is directly related to the amount of design variables considered. VENUS-C proved that cloud computing can dramatically increase that number. The later has an enormous value for the AEC industry as now through this Pilot’s application running on VENUS-C infrastructure it could be feasible to carry out research projects that not so long ago were un-imaginable. (3) Hardening the business model of a Start-up. Including cloud computing applications with proper scalability and interoperability in private and public cloud provider’s results as a robust technological advancement in the business plan. This advancement has a potential value in the eye of the venture capitalist too, in this case interested in negotiating an investment in Green Prefab, with different options in the direct cloud revenue model.</td>
</tr>
<tr>
<td><strong>AEGEAN</strong></td>
<td>Benefit: Civil protection authorities can respond to different workload situations that have to deal with in a 24/7 basis; e.g. unpredictable (fire propagation simulations during a fire event) and/or predictable (daily fire risk estimations) bursting of CPU needs during the months of June-September in a typical wildfire season.</td>
</tr>
<tr>
<td><strong>UNEW</strong></td>
<td>The VENUS-C platform enables performing adaptable and highly scalable QSAR experiments in the Cloud. Unpredictable workloads are handled easily either as a response to users or when new data is released.</td>
</tr>
<tr>
<td><strong>COSBI</strong></td>
<td>Systems Biology research involves handling and interpreting huge data sets. Our cloud-powered solution significantly shortens simulation time compared with workstation executions typically used by researchers and also allows them to address complex challenges that require a higher number of simulations, achieving better results.</td>
</tr>
<tr>
<td><strong>CNR</strong></td>
<td>The D4Science infrastructure offers integrated tools targeted at biodiversity scientists, who need to explore and devise new models based on the analysis of an ever growing amount of data. Unfortunately the productivity of our community is hindered by the long execution time required to perform model validation and refinement. The availability of cost-effective external computational resources is promising, but often hampered by technical difficulties such as application porting between platforms and costs of devising interoperable solutions. We extended the D4Science infrastructure with a cloud interoperability layer that leverages the VENUS-C middleware and eases the porting of existing and future scientific applications in the biodiversity field.</td>
</tr>
<tr>
<td>Project</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
</tr>
<tr>
<td>ImProv</td>
<td>The goal of ImProv is to enable a stable platform for phylogeny computation such that some of the computing nodes are governed by the XtremWeb-CH volunteer computing platform (<a href="http://www.xtremwebch.net">www.xtremwebch.net</a>), some by VENUS-C cloud platform. <strong>This kind of combined high-throughput platform will meet the needs of our main case</strong>, MetaPIGA (<a href="http://www.matapiga.org">www.matapiga.org</a>)</td>
</tr>
<tr>
<td>eRTVMSuC/TRVeC</td>
<td>The VENUS-C framework introduces a new level of scalability into our vessel tracking algorithms. In this way it adds stability and dependability to our Maritime Monitoring capabilities.</td>
</tr>
<tr>
<td>CLOUD-QUAKE</td>
<td>The main achievement realized within VENUS-C for the CLOUD-QUAKE pilot was the development of a dual interactive/automated tool for earthquake shaking simulations using a realistic, stochastic simulation approach, reducing significantly the overall computation time (2 to 3 orders of magnitude).</td>
</tr>
<tr>
<td>Cloud4Trends</td>
<td>Our participation to VENUS-C enabled us to create a robust framework with parallelization capabilities and elasticity features that successfully supports our real-world experiments’ unpredictable workload.</td>
</tr>
<tr>
<td>CAD</td>
<td>VENUS-C and the available platforms have given us the opportunity to access computing resources ‘on-demand’, allowing us to increase our computational throughput and calculate more problems than before.</td>
</tr>
<tr>
<td>Bio-CIRRUS</td>
<td>The pilot we implemented with VENUS-C resources and support allowed us to develop a straight-forward way to run important legacy applications from Bioinformatics in parallel, resulting in important speedups and reduced times and costs. Our test cases in bioinformatics not only succeed in speeding-up the performance and making easier the access to applications but the approach would open a window to address more important challenges in the framework of GWAS and comparative genomics, facilitating data analysis in Bioinformatics with Cloud Computing and Bio-Cirrus.</td>
</tr>
<tr>
<td>cTQm</td>
<td>The main benefit of using VENUS-C is that we can process a large number of transcriptomics experiments and use them to detect possible bias which could not be spotted on an experiment-by-experiment basis. We have shown that there is a clear bias in approximately 15-10% of the experiments which have been publicly deposited.</td>
</tr>
<tr>
<td>CloudERT</td>
<td>The execution of real Monte Carlo treatment verification cases usually needs a large amount of CPU. However, the cost of having the required infrastructure is only viable with a high level of usage. As consequence, Cloud computing, where the computing resources are accessed on demand, constitutes an interesting model. <strong>VENUS-C Cloud platform simplifies the execution of such a process because the COMPs layer manages the provisioning of the computing resources transparently to the User.</strong></td>
</tr>
<tr>
<td>PARSEC</td>
<td>Key benefit: The cosmologists are able to generate (medium-resolution) distribution matter maps in order to confront cosmological theories as well to determine accurately errors.</td>
</tr>
<tr>
<td>TARCLOUD</td>
<td>The robust and powerful VENUS-C framework provided us the means to port microRNA target prediction to the Cloud, making near-real time predictions feasible.</td>
</tr>
</tbody>
</table>

**PILOTS (continued)**
<table>
<thead>
<tr>
<th>Project</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>MoDoC</td>
<td>Using VENUS-C enabled molecular docking solution bio-scientist end-Users can conveniently submit, monitor and retrieve the results of large scale virtual screening experiments. The VENUS-C platform provides a scalable computing infrastructure for the bio-scientists that is available from a convenient and easily deployable User interface.</td>
</tr>
<tr>
<td>CALCIUM</td>
<td>For CALCIUM the main benefit of using the VENUS-C infrastructure is the potential portability of our internally parallel application between different Cloud infrastructures, simplifying the task of moving to a different software stack.</td>
</tr>
<tr>
<td>AOLSBD</td>
<td>VENUS-C proved to us that Cloud Computing can enormously facilitate the study of building’s environmental performance by making feasible, accessible and faster for architects and engineers to perform complex building performance simulations.</td>
</tr>
<tr>
<td>IC-Cloud</td>
<td>ICCloud main goal is to deliver a near real time diagnosis-aided tool for ICU medical practitioners. In order to succeed that we need access in massive storage and computation Cloud infrastructures. <strong>VENUS-C abstracts implementation details, hides complexity and provides a seamless mechanism to diverse Cloud infrastructures.</strong></td>
</tr>
<tr>
<td>ScaBIA</td>
<td>In the ScaBIA we have developed a general method of running Matlab script on the Azure Cloud. <strong>Using this for SPM Matlab scripts for brain image analysis we get the advantages of the cloud with elasticity for computing and storage resources.</strong></td>
</tr>
</tbody>
</table>

5.3 Partners Sustainability and Exploitation Plans
Our partners, most of whom cover a dual role as they are also represent the scenarios that have their scientific application on the VENUS-C platform. Their plans include:
- A list of the VENUS-C components they will exploit
- A statement about the VENUS-C selling factors they identified
- Information on their principal customer base and network

ENGINEERING - INGEGNERIA INFORMATICA SPA
Engineering Ingegneria Informatica spa (ENG) is the leader of the Engineering Group. The Group, thanks to the acquisition of the Italian operation of ATOS Origin in 2009, strengthened its international dimension setting up companies also (other than in Ireland and Belgium) in Latin America (Brazil in 2010; Argentina in 2011) and Middle East (Lebanon in 2012).
Engineering has a multifaceted approach to the market. The various competences in the Group allow Engineering to respond to customer needs, in terms of development projects, products and services to customers in various markets: Public Administrations, Healthcare, Utilities, Defence, Banks and Finance, etc.
Its tier-3 data centre in north of Italy (Pont Saint-Martin, Aosta) and the other two back-up sites in Turin and Padua, allow Engineering to fully support any ICT needs. Recently (in 2010) a Cloud service has been launched to start renewing the current housing and hosting furniture. Such novel commercial offer, based on Cloud services provided an excellent path to exploit in-house infrastructure.

**Principal components of VENUS-C the organization will exploit**
VENUS-C represented for ENG the main R&D investment in open source Cloud technologies in the last period. As coordinator of the project, Engineering managed to promote the overall Company profile and activities at European and international level. As technical partner the ENG team acquired know-how and expertise on several aspects of the Cloud Business.

In particular:

- ENG deeply analysed the potential market and business models for the use of Cloud in the scientific environment and identified a potential market that will be further explored with several initiatives that will be launched in the coming months.
- ENG recognised and understood the different perspectives of the Users and gained insight how to support them, in terms of co-design of services, suitable development and integration of Cloud services within scientific applications.
- ENG developed an (open source) accounting and billing service. The features are based on both the identified requirements of the VENUS-C Users and the available experiences of our Managed Operations team. We will further tests and improve the service with the ambition to integrate them into our CLOE (Cloud of Engineering) services.
- ENG made experiences with OpenNebula v.2.0 and 3.4, gaining know-how on limitations and potential benefits of this open source middleware. The initial benchmark of OpenNebula with other comparable solutions allowed comparing the current status of open source Cloud stack with respect to the business environment. This knowledge will be used to re-design the CLOE services in the next years.
- The whole platform will be made available – as-is and with no warranty – to any interested User, not just VENUS-C Users, as an experimental testbed one year more after the end of the project, to stimulate a subsequent commercial adoption or to allow an easy step-out from those services. Further development or professional services on top of it will be discussed on case by case.

The selling factor of VENUS-C
The VENUS-C platform may serve as the basis for developing scientific application servicing different disciplines. As such it may increase the set of services offered by CLOE specifically targeting the research domain.

Partner’s principal customer base and partner network
The customer base will be based on current internal and research projects where ENG is involved (respectively about 200 and 25). Once the platform is proven to be stable enough it will be exposed as a service through the CLOE offering for its existing customer base in the education and public sector that may fully exploit the VENUS-C platform on a pay-per-use model.

OGF.seeig
Principal components of VENUS-C the organization will exploit

OGF.eeig has led communication and online training in VENUS-C through focused dissemination activities with emphasis on communicating the strategic and operational advantages of the project’s end-User groups, ensuring the web channel acts as a focal point. OGF.eeig has developed the dedicated eTraining environment and modular-based courses on both Cloud Computing in general and VENUS-C Users and services. OGF.eeig has also supported targeted, communication exchange with the new pilots and experiments as part of the activity led by UPV. Practical dissemination guides across the diverse thematic areas and best effort support will facilitate users in continuing to promote their results after the project is completed in May 2012.

OGF.eeig has assessed the feasibility of Cloud Computing for the ‘long tail of science’, small businesses and new services for the public sector. Chief among the objectives has been defining VENUS-C tangible assets for wider exploitation, spanning the broad spectrum of use cases and the technical advances and new services, as well as the eTraining modules. Intangible assets include deep insights into Cloud Computing, interoperability and standards development and implementation, which OGF.eeig has actively supported and promoted, including the role of VENUS-C in the EGI Federated Cloud Task Force and Test Bed. Knowledge of Cloud and standards implementation has fed into the SIENA Roadmap, of which OGF.eeig is a co-author and is informing the EC Study Tender on Cloud Computing for public authorities and eScience, a collaborative effort with IDC.

Specifically, OGF.eeig will exploit VENUS-C overall results through:

- The Cloudscape flagship workshop series (2009-2012), which has been spearheaded by OGF.eeig and is being taken forward as a self-sustainable high-policy annual event in Brussels. Cloudscape IV was attended by 120 stakeholders from the scientific community, industry, standards organisations and policy makers, showing increased industry participation. OGF.eeig shall ensure the use case scenarios and pilots are showcased in 2013, considering there are resources which shall continue into 2013 so it can monitor how VENUS-C has made an impact on their daily eScience activity and report back to the community at large.

- iMarine, which is coordinated by CNR, ensuring the tangible assets of the Aquamaps niche application are exploitable for marine biodiversity as part of the technology and service developments within the project.

- EUBrazilOpenBio, which is coordinated by BSC (with partners from CNR and Universidad Politécnica de Valencia), fostering cooperation between Brazil and Europe in the important area of biodiversity. OGF.eeig will thus contribute to supporting the exploitation of VENUS-C technical developments and applications relevant to the field. OGF.eeig is also leading the socio-economic studies, underpinning long-term sustainability of this international initiative, which is of strategic importance to Europe.

- VENUS-C channel, as part of our responsibility, making the VENUS-C services readily available for public downloads from the dedicated wiki, creating value propositions and leading the final campaign around the open-source release of the components.
• Actively supporting the promotion of the technical achievements for publication in a peer-reviewed journal (e.g. ACM, IEEE, Springer) to validate “consumerisation”.

The selling factor of VENUS-C
The two main selling factors of VENUS-C are the broad spectrum of use cases and the technical advances. The use cases perfectly illustrate the advantages of Cloud Computing not only for the ‘long tail of science’ but also for small businesses and new services for the public sector. VENUS-C is supporting a number of use cases led by small companies, including the Green Prefab spin-out (established by Collaboratorio), Molplex Ltd (a UK start-up), DFRC (a Swiss start-up) and RISC-Software (an Austrian SME). These companies have not only identified new market opportunities in Europe thanks to Cloud Computing but are also driving innovation outside Europe (Asia, U.S.). The availability of VENUS-C components in open source coupled with dedicated value propositions and promotional campaigns will ensure businesses, academic and research institutions can continue to innovate after the project ends.

Partner’s principal customer base and partner network
OGF.eeig has extensive experience in creating global, connected networks in the distributed computing arena, stemming from the organisation’s technical coordination of OGF-Europe and later the SIENA initiative, including the SIENA expert groups (industry group, special liaison group and Roadmap Editorial Board with representatives from global companies, standards groups and European distributed computing infrastructures with an excellent liaison with both NIST and the NSF in the US); the Cloudscape international network; community building in VENUS-C and the new networks being created through the EC Tender Study. Close links have also been established through cooperation work (e.g. StratusLab, EGDI – with 2 use cases as pilots, OpenNebula and liaison with the European Space Agency, among others), standards implementation and social networking activities.

COLB – COLLABORATORIO
COLB will not exploit any of the VENUS-C software components, however will be exploiting their scenario application, Green Prefab HUB-E.

The selling factor of VENUS-C
The “LuxRender in the Cloud” tool in the BIM scenario will provide the Green Prefab User community with a friendly integrated procedure to create an architectural rendered image of prefabricated buildings. Once fully integrated in the Green Prefab system, this service can deliver advantages to an already existing community and can result in a reduction of delivering time in the design phase. This kind of service is sold by competitors (using property grid infrastructures) with a “credit method”, where Users buy credits in advance and then they consume it for each job they submit, or € per core hour. Green Prefab may include the service in a general business plan of the infrastructure and may decide to deliver the service for free to Users, paying Cloud providers with revenues coming from the core business of the platform (% on construction cases).
COLB’s scenario on “Rendering in the Cloud” has been disseminated to all involved stakeholders of the Green Prefab start-up including: architects, engineers, prefabricators and suppliers, building developers, real estate agents, research Institutes, and investors. For the first year after the VENUS-C project the Green Prefab exploitation plan includes the involvement of the community in HUB-E with two separated initiatives:

- “Rendering storms: all community will be enabled for a restricted period of time (2-3 days to be repeated) to use the prototype developed in VENUS-C for free;
- “Manifattura Domani competition”: the new Scientific Park in Rovereto called “Manifattura Domani” signed an agreement (reported in the DOW) to use COLB’s Cloud app in an architectural competition where hundreds of candidates are expected to participate.

Also, in late 2011, Green Prefab Italia srl was founded and pre-incubated in Manifattura Domani, a green innovation factory in Rovereto (Trento). All efforts spent in VENUS-C have been useful to prove to investors that Cloud Computing could be a key-factor for the development of the Green Prefab system in the future. The foundation of the Green Prefab company as spin-off of the VENUS-C project is based on that belief.

Partner’s principal customer base and partner network
The current active networks are in the most promising markets for green building industry: China, USA, Scandinavia, Germany, UK, South Korea, and Italy. There is a growing interest from emerging new economies like Brazil, Africa, and Asia. Green Prefab is going to create local cluster networks in those emerging areas.

Additionally, partnerships with the VENUS-C partner Universidad Politecnica de Valencia and Royal Danish Academy, School of Architecture may enable Green Prefab to offer a first in the world suite of complete Cloud tools for civil engineering.

UNEW - UNIVERSITY OF NEWCASTLE

Principal components of VENUS-C the organization will exploit
A Quantitative Structure-Activity Relationship (QSAR) service for performing automated QSAR modelling. Access to models that have been previously built using the QSAR service on VENUS-C. The QSAR process is composed of a number of steps some of which could be marketed as an individual service. This modularisation will allow people to adapt the process to suit their needs or include new modelling techniques.

The selling factor of VENUS-C
QSAR has many usages in drug development but requires expert knowledge. The automated QSAR service removes the requirement for that expert knowledge and lowers the barriers for SMEs and
academia to compete with large pharmaceutical companies. Using the VENUS-C platform will allow leveraging the resources available in Cloud Computing and provide results quicker and more cost effectively.

**Partner’s principal customer base and partner network**

We have contacts in academia, local SMEs and multinational companies. For instance, Sunderland University has a large pharmacology department with whom we are working to improve their productivity and reduce their costs. Although they are not interested in QSAR directly the components developed are suitable for re-use. Interest from both local SMEs and multinationals (Unilever) is being pursued.

**UPVLC - UNIVERSIDAD POLITÉCNICA DE VALENCIA**

**Principal components of VENUS-C the organization will exploit**

The principal components of VENUS-C the organization will exploit are Bioinformatics Cloud porting, structural analysis Cloud porting, and the pilots’ experiences, feedback and network.

1. **Bioinformatics Cloud porting.** The adaptation model used in this case is quite general and valid for many different bioinformatics applications. By its generalization in a way of plug-ins, it will be possible to ease this process even releasing the components for others to integrate their own tools. We have already started offering access to the tool as a way to reinforce collaborations and to set up new ones.
2. **Structural analysis.** Architrave is a product registered under the UPV and commercially exploited already. The Cloud service constitutes a unique selling opportunity since it enables hosting the core services. Ideal for supporting training courses on high-end construction structures.
3. **Pilots experience.** Through the portfolio of contacts and experiences, we will be able to reach new areas and disciplines, especially nation-wise.

**The selling factor of VENUS-C**

Higher conceptual level for the programming models with respect to IaaS. Higher independency of the framework with respect to commercial PaaS. Capability to deploy a private solution. We think that Users should not be system administrators, but at the same time, locking-in to specific solutions (either commercial or freely available) that PaaS may provide frightens Users in the process of adapting their applications. The use of standards, the support of different Cloud deployment models and the support of companies reduces these threats. Programming models for local clusters and framework should not be much different as programming in the Cloud.

**Partner’s principal customer base and partner network**

Already identified Users’ communities (Bioinformatics and structural analysis). Potential extension to other scientific areas. We have contacts with a wide community of Users in those disciplines, from
previous experiences and through the participation in networking activities. We led the Spanish Network for e-Science, and participated in different distributed computing projects in the past.

KTH - KUNGLIGA TEKNISKA HOEGSKOLAN

Principal components of VENUS-C the organization will exploit
CDMI Proxy. CDMI-Proxy is an implementation of a CDMI-compatible server that can be used to store data both using local infrastructure and public Cloud services.

The selling factor of VENUS-C
CDMI Proxy is a key component for federated Cloud activities and data interoperability. As such, and as being developed within the VENUS-C project, this is a major contribution to the whole Cloud community.

Partner’s principal customer base and partner network
KTH has committed to maintain and further develop CDMI Proxy for public use for – at least – two more years. The Swedish National Infrastructure for Computing (SNIC) has committed to support KTH in this contribution as part of the newly formed SA2 within EGI-Inspire. This SA2 is formed from the EGI Cloud Task Force, where KTH have been contributing from start as a resource provider, and as VENUS-C/KTH for the development and implementation of CDMI Proxy. Through this, the VENUS-C developed CDMI Proxy will be used in a very large network of Cloud Users and provider.

COSBI - UNIVERSITY OF TRENTO CENTRE FOR COMPUTATIONAL AND SYSTEMS BIOLOGY

Principal components of VENUS-C the organization will exploit
Porting and deploying over the Cloud infrastructure services for the description and hence the analysis of complex biological systems.

Porting and deploying over the Cloud infrastructure services for the description and analysis of complex biological systems. The main component used is Generic Worker which is used to enable data flow execution and to run the jobs on Windows Azure. It provides also an STS (Security Token Service) implementation for the User management and security.

COSBI also uses the VENUS-C components COMPSs runtime in order to enable data flow execution and to run the jobs on Linux; PMES-COMPSs to provide the SDK used by the COMPSs client to submit jobs to remote nodes; API .NET SDK for enabling connection to the Generic Worker from the clients.

The selling factor of VENUS-C
• Elasticity, ability to increase or decrease capacity, scaling up and down depending on your needs of performance and space. Pay-per-use, no need to buy costly High Performance Computing cluster for operations which cannot be completed on local PC.
• Powerful scripting environment allows researchers to customize experiments.
• Shorter experiment times result in finer accuracy of results. No additional infrastructure overhead costs.
• No geographical connectivity issues.

**Partner’s principal customer base and partner network**

Interested communities, both academic and industrial, are in the field of medicine, biology and pharmaceutics. We envision particular interest from those labs willing to test the preliminary results of their in vivo experiments against the results of the analysis of cheaper and quicker tests of “in silico” models. Special emphasis goes also on the other side of the coin: tracing emergent behaviours from “in silico” experiments to in vivo experiments.

**BSC - BARCELONA SUPERCOMPUTING CENTER**

**Principal components of VENUS-C the organization will exploit**

In the VENUS-C project, BSC has developed extensions to COMPSs, a parallel programming framework whose runtime has been enriched with advanced Cloud features as requested by the scenarios and extended for interoperability with commercial providers like Azure and not commercial ones using an interoperability layer towards OCCI compliant offerings. Also, the COMPSs Enactment Service, developed from scratch as VENUS-C product, can be exploited in the future in order to provide on demand access to distributed resources.

On the Infrastructure side, EMOTIVE Cloud has evolved to fulfil the requirements acquired from the VENUS-C use cases and pilot applications: OCCI and OVF standards have been adopted to enable Cloud interoperability; the accounting information is now persistent and accessible through an embedded database; certificate-based security has been enabled for allowing VM isolation across the different Users and organizations.

All these components will be leveraged in the EUBrazilOpenBio project for the porting and the execution of the use cases on the cloud infrastructure.

**The selling factor of VENUS-C**

The previously described extensions will allow BSC to offer a programming framework as a complete platform for distributed environments, reaching a wider number of User communities. The synergies developed during the project have brought the COMPSs framework to a more mature level allowing its adoption not only as a research product used in the BSC supercomputing facilities but also as a component of more complex infrastructures supporting industry.

**Partner’s principal customer base and partner network**
Although BSC has industrial collaborations with a large number of organizations (such as IBM, Microsoft, Repsol, SUN, NVIDIA), in this section, only the collaborations and links related to Distributed computing are listed. Most of distributed computing related collaborations with industry are through projects funded by the EC. BSC has participated in the following projects: BEinGRID, BREIN, SORMA, XtreamOS and IS-ENES, all of them with a high participation of industry (like Telefonica I+D, ANSYS, ATOS Origin, MICROSOFT, IT Innovation, Elsag Datamat, BT, CETIC, SAP, EADS, XLAB, and many others). BSC is part of LA Grid initiative, project sponsored by IBM. In this project, BSC is contributing to two specific research projects related to Grid computing. BSC is a member of NESSI. BSC is also a partner in the recently funded Spanish level Cloud project NUBA, coordinated by Telefonica I+D with other industrial partners like ATOS and SMEs like Caton, Digital Bubble and Xeridia.

CNR - CONSIGLIO NAZIONALE DELLE RICERCHE

Principal components of VENUS-C the organization will exploit

The VENUS-C support for workflows management, for batch execution of large data challenges, and the event stream processing support are the application paradigms candidate to enhance the technology current exploited by CNR in the D4Science infrastructure. In particular, the following capability will be exploited within the project timeframe: execution of workflows that compose specific registered or added on-demand on the Cloud applications to elaborate scientific datasets that need to be retrieved, partitioned, elaborated, annotated, and stored on existing storage elements. Tools, services, and APIs to register and unregister resources, to monitor their usage, and to account their exploitation performed by clients are the required add-ons to make feasible the integration of VENUS-C facilities with the existing D4Science e-Infrastructures.

Up to now, CNR focused on the exploitation of components belonging to both of the two main groups supplied by VENUS-C:

- Programming models enactment services
- Support functionality

The programming model enactment services are the entry points for spawning computation in the Cloud. They share the same goal, namely execute researcher’s code in the Cloud, while offering a different set of primitives and characteristics. The Generic Worker programming model offers a more low level approach, requiring the User to directly orchestrate the job submission in order to parallelise the code, whereas COMPS superscalar (COMPs) offers an integrated environment which helps build complex parallel applications by transparently spawning child jobs and managing their lifecycle and dependencies. Developers who port their software directly to VENUS-C would select the programming model that fits best their application. We decided to build a generic layer that will enable the running of the User’s code in both environments. The main reason for this is that the two programming models have each a different target platform (COMPs is Linux based, while the Generic Worker is Windows based). Another benefit of the choice was to allow testing of both platforms. Despite that the latest
version of COMPSs can be used to submit jobs to Azure, the support of GW was kept since it provides additional advantages (e.g. no need of an additional endpoint running COMPSs enactment services).

The **support functionality**, on the other hand, consists of shared components between platforms; among those the following have been exploited.

The **CDMI proxy** is used in all cases where it is necessary to interact with the Cloud blob storage, such as updating the application executables used by the Cloud executor or when the direct access to the RDBMS was impossible due to firewall restrictions. CDMI is also used to manage the deployment and updating of applications, also managed transparently through RainyCloud.

The **Accounting & Billing** service is used since we need a reliable way of keeping track of resource consumption at a fine-grained level. Finally, the **Traffic Redundancy Elimination** (TRE) component is used to reduce the time and for frequently transferred data between the Cloud and the infrastructure on premises.

**The selling factor of VENUS-C**
Activities such as fisheries, tourism and recreation, and shipping are largely affected by the wealth of oceans and large marine areas. Predicting the impact of climate changes on those ecosystems is the way to protect the goods and services they provide and the economic systems that depend on them.

To such end, our work has aimed to demonstrate that Cloud Computing leads to a significant reduction in the time needed for performing bio-climate analysis on marine species. The generation of a probability model for species is a highly computational intensive process, which is parallelisable in most cases. Our analysis shows that it is currently possible to extract non-trivial behavioural phenomena about species distributions in the oceans.

Parallel computing has enabled us to perform – for example – probability projections reducing the time from 7 hours to 38 minutes. In other cases – to evaluate probability variations – a local single threaded computation would have required 35 hours; using a parallel computation, the selected algorithm required 3 hours, resulting in an 11x speed-up.

A Cloud-based solution enables us to leverage parallel computation in a more flexible and cost effective way than grid computing or high performance computing solutions on premises.

**Future developments**
Evaluating the impact of storms, tsunami, and other natural disasters on local marine ecosystems and in turn on economic systems is fundamental to response, mitigate, and recovery in a prompt and efficient way the damaged area. Prediction and evaluation of the impact of climate changes and marine disasters will be investigated, being one of the mandates of major international organizations such as the FAO.
Partner’s principal customer base and partner network

Although CNR has collaborations with an extremely large number of scientific and industrial organizations, in this section only the collaborations and relationships related to distributed computing are reported.

Most of those collaborations are established through projects funded by the EC. CNR participated or is participating in the following projects in the area: DILIGENT, D4Science and D4Science II, DRIVER and DRIVER II, OpenAIRE and OpenAIREplus, EFG, HOPE, ENVRI and iMarine; all of them with a high participation of notable research institutions (e.g. CERN, CNRS, European Space Agency, FAO, Fraunhofer Institute, Max Plank Institute, STICHTING SURF, WorldFish Center, Universities of Athens, Bielefeld, Edinburgh, Glasgow, London, Nottingham, Warsaw, etc.).

From a scientific point of view, the disciplines covered by these activities range from High Energy Physics to Biodiversity, Aquaculture, Earth Observation, Cultural Heritage, etc. The spectrum of beneficiaries/customers is also extremely wide, ranging from Digital Libraries developers to Statistical Data Services providers, Geographic Information Systems providers, Niche Modeling developers, etc., up to complex hybrid e-Infrastructures for e-Science.

Microsoft (MRL- MICROSOFT RESEARCH LIMITED, EMIC- EUROPÆISCHES MICROSOFT INNOVATIONS CENTER GMBH, MICROSOFT INNOVATION CENTER GREECE)

Principal components of VENUS-C the organization will exploit

In the first place, we intend to exploit the GW platform components that have been developed in the project. This will be done by applying a twofold approach as follows:

- As indicated in the description of work, Microsoft will continue to provide direct administrative and technical support for the pilots and experiments acquired through the VENUS-C open call for another year. This support will also cover the VENUS-C scenario partners and consists of free access to Azure resources, technical training, help desk support as well as bug fixing and maintenance of the released GW software packages. We are excited to announce that a full time head has been secured and already assigned for this task.

- Beyond this support of the existing community, Microsoft intends to enhance the User community by actively approaching the other participants in the VENUS-C open call that could not be considered by the project, but also other European academic partners of Microsoft. For the latter, we intend to also leverage Microsoft’s Azure for research engagement project, see http://research.microsoft.com/en-us/projects/azure/ where first VENUS-C results have already been disseminated e.g. at the Cloud Futures workshops 2011 and 2012. The idea is to not only further disseminate the VENUS-C platform and foster adoption, but also to spread results coming from the engagement projects among the VENUS-C Users. We plan to involve two other VENUS-C partners and have already started discussion how to share support of the general management of the community by means of joint conference calls, the organization/support in
the organization of dedicated community events and dissemination on the web and social media. For the execution of the plans, it is anticipated to allocate a dedicated budget and team.

From all the activities mentioned above, feedback and requirements will be gathered and shared with the research community to enable development and support of new features of the GW platform components which will be released as open source. As an additional result we hope to be able to develop and provide new Cloud and/or Azure capabilities to meet the specific needs of scientists and governments with a view to e.g. interoperability or regulatory aspects. We plan to prepare public case studies and to share the findings with the European Commission and industry.

Ultimately, this could result in new project ideas and partnerships in time for Horizon2020. The overall ambition is to further position Cloud Computing as a viable instrument for researchers and the funding agencies in support of a European hybrid Cloud model that covers both public and commercial offerings.

Microsoft, through the deployment of the Azure platform in Europe, has started to gain experiences in fulfilling the requirements and supporting demanding User communities such as the scientific and governmental ones. In general, such communities are a good experimental base for further commercial deployment, and will be giving insight to better understanding whether and how well they can be served by Cloud offerings. Through timely close collaboration with Microsoft’s developer and evangelism unit in Europe we intend to further sustain the activities and lay foundation for a later (commercial) uptake in the field.

**The selling factor of VENUS-C**
The VENUS-C PaaS on top of Azure is one of the key selling factors of VENUS-C as it will provide a thin abstraction layer for applications that do not need to understand the underlying hardware. In this way, VENUS-C will promote interoperability among different underlying solutions.

**Partner’s principal customer base and partner network**
Beyond the dedicated communities mentioned above, Microsoft has a huge customer base of all different groups (businesses, professionals, home Users) and at different software levels (operating system, daily software tools and applications) and closely works with a huge partner network with country subsidiaries and different levels of collaborating partners.

**Principal components of VENUS-C the organization will exploit**
TRE, also known in the industry as “WAN accelerator”, enables a significant speed-up of the communication for the client due to the traffic reduction induced by the TRE operation.

**The selling factor of VENUS-C**
TRE delivers to VENUS-C end-Users a novel receiver-based end-to-end solution that relies on the power of predictions to eliminate redundant traffic. In this solution, each receiver observes the incoming stream and tries to match its chunks with a previously received chunk chain or a chunk chain of a local file. Using the long-term chunks meta-data information kept locally, the receiver sends to the server predictions that include chunk signatures and easy to verify hints of the sender’s future data. Upon a hint match the sender triggers the TRE operation saving the Cloud’s TRE computational effort in the absence of traffic redundancy.

**Partner's principal customer base and partner network**

IIT will continue to serve as the TRE focal point of VENUS-C partners, assisting them in adopting and designing the right TRE scheme. We will continue to provide information and consulting in this area.

---

**AEGEAN – UNIVERSITY OF AEGEAN**

**Principal components of VENUS-C the organization will exploit**

Within VENUS-C, AEGEAN uses the Generic Worker (GW) programming model with Windows Azure to develop its Emergency and Civil Protection scenario. Cloud provides reliable data storage and scalability of the workload computations. Furthermore, AEGEAN uses the components of security, notification scaling and accounting provided by the GW programming model. Besides the use of Azure, AEGEAN runs job execution in the Cloud infrastructure (Virtual Machines with W2008R2) provided by ENG. This demonstrates the ability of the system to run into multiple Cloud platforms and contribute to interoperability.

**The selling factor of VENUS-C**

Efficiency of the calculations for the fire risk and fire propagation is ensured by reducing the time needed for the computations and by serving simultaneously multiple Users. Cloud computing provides reliable data storage and scalability of the workload computations, including increased spatial and temporal resolutions. The instant and prompt availability of processing power along with the cost effectiveness of the Cloud-based solutions are key elements that are better suited to our scientific problem at stake. For non IT end-Users, the Cloud solution may be more User friendly than supercomputing in which considerable technical experience may be required. Regarding the cost of the Cloud solution, it is more efficient to use VENUS-C services rather than purchasing new hardware. Fire risk and propagation applications are useful during the high risk season that can be extended up to six months. Therefore, new purchase of hardware (needed to support an ever increasing User and computational load) will probably remain unexploited for almost half of the year; furthermore, maintenance and administration costs will be significantly high compared to the ones of VENUS-C costs. Wildfire management authorities prefer to invest their resources to fire fighting equipment and services rather than in less-necessary on-premise computational resources.

**Partner’s principal customer base and partner network**
AEGEAN’s partner network includes the end-Users of the proposed system. The network includes civil protection authorities and managers, fire fighting forces, emergency crews, forest managers and wildfire specialists (serving operational, academic and/or research purposes). However, model developers can also be assisted in a conceptual architectural design and development of similar wildfire management systems based on Cloud Computing technology for different areas and in a largest spatial context (e.g. regions and countries). AEGEAN has already presented the system to 2 groups of fire fighting officers (in Thessaloniki and Mytilene, Greece). Finally, the Cloud fire risk algorithm may be used in another research project (XENIOS) as a validation/dissemination activity for VENUS-C.

6. CONCLUSIONS

Our study began by examining and considering the different sustainability perspectives of Scientific Clouds and how to judge the usefulness to its users. The process brought us to research the current state of the Cloud landscape (specifically the different Cloud business and deployment models), its potential, the adoption aspects and their suitability in regards to scientific users.

This analysis took a closer look at the Cloud model through the experiences, usage and behaviour of our User community. Based on the analysis of the findings obtained through our VENUS-C sustainability questionnaire and relevant research a careful investigation of the benefits was done. In addition, a SWOT analysis was conducted, identifying the gaps and potential opportunities of adopting Cloud Computing in the eScience. Taking into account all these as input a sustainability strategy was formulated, which also researched the key associated elements: technical, organisational, policy, legal and financial.

VENUS-C has produced significant tangible and intangible exploitable assets, of which include 6 open source software components, 7 scenario applications, a great advancement of Cloud Computing knowledge and contribution to standard bodies, experience and know-how. Our community (partners and users) have much to exploit. All partners who are involved in providing the platform and services (namely Microsoft, BSC, ENG and KTH) will continue to do so for an additional year on a best effort basis.

That certain features of Cloud Computing are attractive and useful to the scientific community, is not questioned by any of the stakeholders involved. But, when exploring whether there exists a true potential for Scientific Clouds to be sustainable and key for the scientific community, we found that there were many key factors that affect sustainability, all of which needed to be taken into account.

The success of VENUS-C was the user-driven design and more precisely the success in fulfilling the actual user community requirements, which has clearly demonstrated the usefulness of the Cloud for our user community. We can conclude that the Cloud Computing model is useful for the communities and disciplines (11 different disciplines, from 10 different Member States) represented by the VENUS-C user community. Furthermore, their (both SME’s and academic institutions) actual usage behaviour (workload type, nature of jobs, resource sharing level), coupled with their “wants and needs” make Cloud Computing a good fit.
Our Community Key Findings

1. Specific, tangible benefits gained depend on individual implementations of the cloud concepts.
2. Cost savings are not at the forefront of researchers’ mind when deciding to use the cloud, rather they are driven by better performance or a new capability.
3. The Cloud for eScience equals “better science”, as it increases quality for Users as it provides benefits related to new tools and resources.
4. Cost effective usage behaviours could include (1) unknown and (2) sporadic demand.
5. Our community real usage patterns are (1) Sporadic peak usage, (2) Oscillatory demand, (3) Plateau of resources, all of which have been embraced in VENUS-C and have perceived advantages of Cloud Computing.
6. Identification of potential barriers, which includes: vendor lock-in, legal and financial issues, training.

Therefore, VENUS-C is a confirmation of advantages of Cloud Computing not only for the ‘long tail of science’ but also for small businesses and new services for the public sector. Our use cases perfectly illustrate this and confirm that VENUS-C offers a great place to do great science.

Our sustainability questionnaire respondents were asked, What would be the best Cloud model for your application in the near future? They were asked to rank each deployment type (private, community public and hybrid), on a scale from one to five. The hybrid model got the highest percentage at the highest rank, therefore the highest score. The common thread of those respondents that preferred the hybrid Cloud was, according to them, the benefit of accessing high computing power, but with restrictions and legal issues/dealing with sensitive data being better cared for. Other respondents notes that (1) the hybrid model could see fewer issues to convince to potential collaborators/research and (2) It would be the most effective, as it can deal with low-scale experimentation and reuse local infrastructures. We agree with our respondents, in that the hybrid approach would likely be the best bet for a good steady start to the possible destination being public Cloud. We agree with our respondents, in that the hybrid approach would likely be the best bet for a good steady start in Cloud Computing.

Currently, the move to a public Cloud could very well be a considerable decision for scientific Users to take, despite of uncertainties related to data control (distribution etc.), legislation and others. However, a hybrid approach would immediately cater to these needs, while in the background other research and development would likely continue to close these gaps and challenges and make way to take on a larger market, covering also scientific Users who have particular needs and worries and would largely benefit from the issues related to control versus cost. The private/community Cloud can thus be serving the needs of the big and well-organised communities with complex requirements, while the commercial public cloud resources will serve the long tail of scientists and possibly some of the cloudbursting needs of the first. Therefore, we determined the hybrid Cloud model (community cloud + commercial cloud resources) for research to be a promising sustainability approach for e-Science Clouds.
7. REFERENCES & RESOURCES

All web references were accessed before 29/May/2012.


Gartner, “Hype Cycle for Cloud Computing”, ID number G00214915, July 2011


http://accelrys.com/

http://en.wikipedia.org/wiki/Long_Tail


http://www.btplc.com/
ANNEX A - SUSTAINABILITY QUESTIONNAIRE ANALYSIS

Below is the list of the respondents, in submission order:

<table>
<thead>
<tr>
<th>Project/Scenario Name</th>
<th>Entity Name</th>
<th>Respondent Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>CloudERT</td>
<td>CESGA</td>
<td>Carlos Mouriño</td>
</tr>
<tr>
<td>Bio-CIRRUS</td>
<td>University of Malaga</td>
<td>Oswaldo Trelles</td>
</tr>
<tr>
<td>TVEReC</td>
<td>DFRC AG</td>
<td>Erel Rosenberg</td>
</tr>
<tr>
<td>TARCLOUD</td>
<td>ATHENA RC / IMIS</td>
<td>Theodore Dalamagas</td>
</tr>
<tr>
<td>MoDoC</td>
<td>University of Westminster</td>
<td>Tamas Kiss</td>
</tr>
<tr>
<td>CALCIUM</td>
<td>HERIOT - WAIT UNIVERSITY</td>
<td>Idris skloul Ibrahim</td>
</tr>
<tr>
<td>IC-CLOUD</td>
<td>University of Cyprus</td>
<td>Nikolas Stylianides</td>
</tr>
<tr>
<td>Scenario - AquaMaps</td>
<td>CNR</td>
<td>Pasquale Pagano</td>
</tr>
<tr>
<td>Scenario: Bioinformatics/Structural Analysis</td>
<td>UPVLC</td>
<td>Ignacio Blanquer</td>
</tr>
<tr>
<td>CLOUD-QUAKE</td>
<td>RESEARCH COMMITTEE Aristotle University of Thessaloniki</td>
<td>Costas Papazachos</td>
</tr>
<tr>
<td>Scenario - Systems Biology</td>
<td>COSBI</td>
<td>Angela Sanger</td>
</tr>
<tr>
<td>CAD</td>
<td>Molplex Ltd</td>
<td>Vladimir Sykora</td>
</tr>
<tr>
<td>Scenario - Building Structural Analysis</td>
<td>UPVLC</td>
<td>Jose M. Alonso</td>
</tr>
<tr>
<td>Scenario - Drug Discovery</td>
<td>Newcastle University</td>
<td>Simon Woodman</td>
</tr>
<tr>
<td>ScaBIA</td>
<td>KAROLINSKA INSTITUTETDEPT OF WOMAN’S AND CHILDREN’</td>
<td>Scalable Brain Image Analysis</td>
</tr>
<tr>
<td>AOLSBD</td>
<td>Royal Danish Academy of Fine Arts</td>
<td>Emanuele Naboni</td>
</tr>
<tr>
<td>PARSEC</td>
<td>CIEMAT</td>
<td>Miguel Cárdenas-Montes</td>
</tr>
<tr>
<td>Scenario - Civil Protection &amp; Emergencies - Wild Fire</td>
<td>University of the Aegean</td>
<td>Prof. Kostas Kalabokidis</td>
</tr>
<tr>
<td>Cloud4Trends</td>
<td>RESEARCH COMMITTEE Aristotle University of Thessaloniki</td>
<td>Athena Vakali</td>
</tr>
<tr>
<td>BIM Building Information Management - Green Prefab GPF</td>
<td>COLB Collaboratorio</td>
<td>Furio Barzoni</td>
</tr>
</tbody>
</table>

Table 7 – List of Questionnaire Respondents

Below is the full questionnaire and analysis

Usage Behaviour
1. How do you manage the data for your application? Please check all that apply.
   a. Use and expect a parallel file system
   b. Use and expect local disk
   c. Large amounts of static data is required at the site where the application runs
   d. Input data arrives from offsite
   e. Output data needs to be pushed out to a different location

The majority (45-60%) of the respondents selected that their output data needs to be pushed out to a different location, use and expect local disk, large amounts of static data is required on site where the application runs, and input data arrives from offsite.

(graph found in the document)
2. Provide rough orders of magnitude on running time of a single application run. (graph and findings found within the document)

3. How many runs do you typically perform? And how frequently? (graph and findings found within the document)

Technical Aspects

1. What were the biggest barriers you encountered when getting started with the Cloud?

Below is a list of the different barriers encountered. The list is lengthy, but worth presenting, same for the benefits perceived listed as a separate question below.

- The movement of data to from the Cloud.
- Time elapsed during provision
- Our cluster required too much technical support and maintenance and the overall reliability was not sufficient
- High learning curve of the programming model
- Instability and incompleteness of the environment. Several bugs have been reported that significantly delayed and slowed the work down
- Infrastructure, and software compatibility
- To understand the execution model
- Need to study the APIs and the protocols to use
- Provision to resources is still very low-level and requires cluster administration skills
- Preconfigured images work well in small-scale cases
- Adapting to the programming requirements of the specific Cloud infrastructure
- Adapting to the concept realized by the GW module, especially the fact that it does not allow multiple runs on a single Cloud node (multi-threading, multitasking)
- Choosing the best programming model that fits the Cloud infrastructure and solves our research problems (e.g. data flow versus work flow)
- Working with non-relational databases
- Dealing with systems that are less stable and are not transactional
- Setting up a local database to store the inputs of our application, then launching jobs remotely.
- To learn the development environment and the differences in architecture and development of the Cloud approach with regard to the traditional approach.
- To migrate the local application to the Cloud
- To prepare and set ready to use the execution environment
- Customization of Cloud platform
- Finding technical support- crucial for those that do not have an in-house developer or programming experience
- The completeness and the clarity of the documentation was lower than desired in order to be autonomous in the porting process
- The application porting was not as straightforward as expected
- The elimination of dependencies from third parties vendors (e.g. ESRI’s ArcGIS scheme)- this demanded re-programming of all the calculation algorithms
2. Do you pay any software licenses?
   2 respondents pay for licenses (Windows an Intel Compilers, Visual Studio 2010)

3. Do you pay any software subscriptions?
   All said no.

4. In regards to Job Execution.
   How would you rate the *elasticity* of the service provided? *Elasticity* being the ability to increase or decrease capacity, scaling up and down depending on your performance needs.

   More than 75% of the respondents ranked between 3 and 4

5. In regards to Data Management.
   How would you rate the *elasticity* of the service provided? *Elasticity* being the ability to increase or decrease capacity, scaling up and down depending on your needs of performance and space.

   Over 60% of the respondents ranked 3. None ranked 1 and 2.
6. In regards to Data Management.

How would you rate the reliability of the service provided? For example the possibility to recover. Over 65% of the respondents ranked 3 or 4.
7. In regards to Accounting and Billing.  
   How would you rate the *ease of use* of the service provided?  
   55% of the respondents ranked 3 or 4.

8. In regards to Accounting and Billing.  
   How would you rate the *flexibility* of the service provided?  
   Over 72% of the respondents ranked 3 or 4.
9. How would you rank your experience (in terms of difficulty) with workflow management in a virtualized Cloud environment? Low (Easy), Medium, or High (Difficult)

Exactly 61% of the respondents ranked their experience with workflow management in virtualized Cloud environment as Medium. Of the remaining 30%, the majority said High.

![Figure 24 – Questionnaire Answer – Ranking of experience with workflow management in a virtualized Cloud environment](image)

10. Do you feel that as part of your involvement with VENUS-C you had the tools to effectively manage and operate virtualized Cloud environments?

Only 1 respondent said no. Amongst the positive answers, some comments included: all tools were there, learning curve high or needed previous experience, but mainly that set of functionalities was high enough in order to fully manage the virtualized environment.

11. Have you encountered performance challenges? For example: slow I/O, network congestion...

45% of the respondents noted NO performance challenges, of those that did, a commonality was slow virtual Machine CPUs core, slow disk I/O and network congestion due to the large volume of data to be retrieved-- all of which are accepted by the Users, as the benefits they perceived out weight these downfalls (see below related question 14).
12. Have you encountered reliability challenges? For example: loss of data, data retrieval issues...

Over 65% noted NO reliability challenges, of those that did, a commonality was issues with stability of the COMPS's, and GW but these and others have made further progress since the questionnaire was distributed.

13. Have you encountered scalability challenges?

20% of the respondents were not in condition (use of infrastructure) to fully answer the question, of the remaining 80%, over 90% perceived NO challenges.

14. Virtual machines are a popular mode of operation in Cloud Computing that tends to affect some applications in terms of performance. Does this affect your application? Please specify.

Approximately 40% of the respondents said that virtual machines do NOT affect their application, while of the remaining 60%, approximately 70% said YES, but the other Cloud features are more attractive.

Financial/Funding Aspects

1. Would you be able to find funding for operating in a public Cloud?

Approximately 50% of the respondents said YES, of the remaining 50%, half of the respondents answered with NO and the others said MAYBE.

2. Would you consider partially funding your work personally?

39% of the respondents DO NOT consider funding their work. Of the remaining 61%, 27% were UNDECIDED, and 72% said YES. Of those whom answered YES, 2 respondents gave a figure of how they would be willing to spend per month- an average of 110 euros/month.

3. Can Cloud resources expenses be considered as an eligible cost in the competitive calls you participate?

67% of the respondents answered YES. Although, some respondent’s comments included that it was eligible “somehow”, but no explicit statement in the calls.

Overall

1. How much experience did you have with Cloud technologies before the grant?

Rank from 1 to 5, five being the highest score

Approximately 1/4 of the respondents ranked 1, being the lowest score.

Of the remaining respondents, the majority ranked 3 or 4.
2. Describe your situation before the grant. Why did you decide that Cloud could be useful?

Below is a brief summary (often quoted directly from the answers) of all the respondents answers- italics why they decided the Cloud could be useful.

- We developed and tested an applications running in local clusters and Grid and we need more resources for put the application in production and due scalability issues
- We work for the National Institute of Bioinformatics in Spain. Cloud computing is an alternative for offering services. We are evaluating this alternative
- We were in a transition toward migration of our application to Cloud.
- Due to our involvement with scientific datasets that need real time processing, we decided that Cloud-based data management solutions would be the right choice.
- We have been involved in nationally and EC funded Cloud-related projects in the past few years investigating mainly Eucalyptus and OpenStack based Clouds. The motivation is that the sustainability of our research infrastructure is not clear. It is uncertain whether new funding for clusters will be available or not. We consider Clouds and desktop grid computing as the two viable alternatives.
- A PhD student, working in Computer Networks. Yes, I felt that Cloud could be a good jump in Computer resources use, and would be very beneficial, especially for those who pay a lot of money for getting a new, or faster hardware
- We have an application that retrieves physiological parameters from ICU patients. Our data was locally stored thus could not be shared easily. Moreover, we could not run intense algorithms to identify clinically interesting episodes; therefore we looked to the Cloud.
- The CNR team joining this initiative operates a private Cloud with 250 virtual machines to serve different scientific communities. This private Cloud is growing constantly since 7 years and it is evident that it is not our primary goal to maintain and operate a data centre. We decided to invest in Cloud technology in order to be ready when it will become possible for us to use public Clouds.
- It was clearly a model being identified as a potential source for computing and data resources. The provisioning model fits very well the working behaviour. We had previous experiments on Grids that lack from sustainability plans, which were de facto provided in Clouds by the use of public infrastructures.
- Conducting computation systems biology research and algorithmic modelling can require heavy computational resources. Not all researchers, companies or academic institutions have access to these.
resources (e.g. HPC) making the Cloud an attractive solution especially since it is offers scalability (elasticity), pay-on-demand/pay-as-you-go, no additional IT overhead costs and no geographical connectivity issues.

- We are software-oriented company and need to perform large computations. **Cloud gives us the flexibility to have computing resources easily scaled as needed.**

- Before Venus-C, we have deployed a Structural analysis Grid Service. Notwithstanding, we decided to use the Cloud due to the advantages that this technology offers faced to the Grid technology, i.e., high availability of the platform, pay-per-use, elasticity, no need of purchasing and managing hardware resources, etc.

- It was interesting to research and to show the capabilities of Cloud Computing for medical image analysis.

- Before the grant we knew that Cloud Computing existed but it was an utopia to think that we could use it and that it was possible to use it with our applications.

- We thought it was useful as it could improve our research and professional activities by considerably improving the execution time of building performance simulation.

- We were very interested in testing this paradigm by using one of our applications. This allows us to check the capabilities of the Cloud to manage our scientific problems.

- The current situation in Greece is based on low demand computation algorithms of fire risk with low spatial resolution. Research efforts of the University of the Aegean (AEGEAN) prior to the VENUS-C project included the development of a fire risk algorithm based on a quantitative assessment of fire danger in a systematic way.

In addition, **our Virtual Fire system had two shortcomings that intended to be resolved by migrating to the VENUS-C platform.** First, to conduct fire behaviour simulations, an authorized User added a new fire ignition point from the application’s GUI. Then, a system administrator conducted the fire behaviour simulations on a local computer by providing the User inputs. Users had to wait for the simulation results that were uploaded on the system through a multi-step complicated procedure. This method was neither automated nor efficient and fast. Furthermore, if a system administrator was not able to process the User’s request (e.g. end of business day) Users could not receive any results. This task was completely “human-dependent”. The second shortcoming of the system was the calculations conducted for fire risk mapping. The system was able to produce just one map per day for a relative small area (Lesvos Island) through calculations based on High Performance Computing (HPC). Even though it produced daily fire risk maps (in a deployment with two HPC nodes), hourly-resolution predictions could not be produced due to processing and hardware limitations.

All of the aforementioned **limitations have been successfully faced** by porting the application into the VENUS-C infrastructure. **VENUS-C provides not only reliable data storage but also scalability of the workload computations. Efficiency of the calculations is ensured by reducing the time needed for the computations.**

3. Does the current Cloud service (Azure and/or ENG/KTH/BSC) satisfy your needs?
   Just over 65% of the respondents said YES. The respondents that said NO cited the reasons being that they hadn’t used the infrastructures extensively yet to be able to give an evaluation.

4. What are the benefits you have perceived from the Cloud service?
Below is a list of the different benefits perceived:

- Elasticity (allocated CPU power needed)
- Ability to assign the right amount of resources
- Access to large amounts of resources/processors
- Managed to improve its execution time by re-designing its workflow, identifying processes that can be run in parallel, and exploiting the Cloud service to execute them
- Easy to migrate application
- Provide application through a simple custom interface for Users
- Assessing sequential programs vs. parallel ones
- Data distribution and safety
- Parallelization in real time
- Ease of use
- Reduced time to generate data (execution time reduced)
- Use of standards: client application can be executed from any OS and can submit jobs to different providers
- The enactment of data-flow jobs (GW and COMPSs)
- The possibility of metadata (CDMI)
- Deployment is easy
- Interoperability thanks to the tools to execute jobs over different platforms.
- Ease of management of resources.
- Pay-as-you-go for large resources
- Reducing cost of ownership of hardware (lower investment cost in workstations)
- Possibility to choose between different platforms
- Possibility to execute complex simulations
- Unattended long execution
- Economic benefits can be invested in other (core) needs rather than IT requirements

5. Do you feel the Cloud is technically sound for your scientific community and the application you use?

Approximately 85% of the respondents said YES, the remaining said STILL NOT. Those that said STILL NOT noted the reasons: need for more tools for security and tracing of the data, and API to control scaling. (graph found within the document)

Below is a complete list of all the respondents’ elaborated answers.

- Yes, We could exploit our service in a pay-per-use way
- Still not. More tools for security and tracing of the data are needed. Biomedical data is highly sensitive
- Yes it is. The main problem with Cloud today is the need to outsource your data, as for scientific community this is minor issue therefore Cloud is optimal for such applications.
- Yes, if the bugs and issues will be fixed and the platform becomes more stable, then yes. The Cloud platform is completely suitable for our application that requires significant CPU but not particularly data intensive.
- Yes. I felt that I’m going to be a Cloud personal, in its technology and use. As an academic person, I will carry on the Cloud research side.
• Yes. especially in medical applications
• YES, it is. Reducing the time for an execution it is possible to submit and elaborate different models in a reasonable time. Scientists are in the position to make progresses in their science faster and better.
• Yes, although it requires a bit of "robustisation" that is aimed for the following months.
• YES, I do not see any reason to contradict this statement. Although we have not incorporated complicated computing processes (like running MATLAB applications, etc.) we have successfully run standard (Fortran-C) executables, a typical product of scientific community research work.
• Yes. Usually researchers need simulate large amounts of data; causing potentially a huge amount of data to download; however, our implemented solution gives the researchers the ability to automatically select only relevant data for downloading (which is significantly less than the actual output).
• Yes. It provides the interface to launch jobs and scale up/down resources. Although an API to control the scaling would be beneficial.
• Thanks to the Cloud service implemented, researchers will have available a huge number of computational resources to be on-demand employed and lots of cost-effective simulations will be launched simultaneously. Thus, more structural experiments will be analysed per time unit, increasing the number of structures simulated and speeding up the research process. If researchers get more results in less time, they can get more and better conclusions in their research and reducing the time-to-publication. The structural community will be able to solve larger scale problems, increase the complexity of the structure to be analysed, and carry out a larger number of realistic dynamic simulations without simplifications. In this way, the reliability and safety of the results obtained will be improved and new structural problems will be tackled. Since the time spent on the design of buildings and civil engineering structures will be reduced, the engineering companies and the architectural studios will increase easily their productivity and volume of business. Finally, there will be no need of acquiring software licenses in property and expensive hardware for solving large-scale structural problems (just pay per use), and the Users will not be worried about new software updates.
• Yes
• Yes, however security issues regarding data protection is needed to be solved. Yes, but the barriers exposed in the previous topic of this survey have to be mitigated. Astronomy and Astrophysics are suffering a deluge of data, so that this community is changing the protocols, adopting digital infrastructures as a fundamental part of their daily work. This process will not have a return point. Cloud Computing perfectly fits to the Civil Protection and Emergencies application, because of the large amount of data and the complexity of the processing. Cloud Computing provides reliable data storage and scalability of the workload computations, including increased spatial and temporal resolutions. Efficiency of the calculations is ensured by reducing the time needed for the computations. More specifically in wildfire prevention and management with Cloud Computing, authorities can respond to different workload situations that have to deal with in a day-to-day situations; e.g. unpredictable (fire propagation simulations during a fire event) and/or predictable (daily fire risk estimations) bursting of CPU needs during the months of June to September in the fire season of the Mediterranean-type of ecosystems.

6. Is there anything you would have liked to have done with the Cloud service but could not?

50% of the respondents said YES, and the other 50% said NO.
Below is a list with the elaborated answers (often directly quoted) of those respondents who answered YES.

- Perhaps some of the applications I have ported to Grid
- Security
- Storage and processing of sensitive data
- I would like to be able to use for free a database or be able to install in VMs my own software.
- The current GW module does not allow multitasking-multithreading. However, we have realized this with standard Azure tools.
- The ability to specify the worker instance to be removed from the platform, to protect appropriately the privacy of Users and applications.
- Update VM images without re-transferring ALL the VM
- Flexibility of deploying custom application dependencies over different Cloud vendors
- Would like to test very long unattended execution (more than 1 month).

7. What do you feel are the major obstacle(s) you have to overcome to be able to use a public Cloud? For example: Legal, financial issues …

Below is a list of what obstacles the respondents personally feel they need to overcome to be able to use a public Cloud. Notable, financial issues were the most reoccurring obstacle.

- Legal issues about licensing and data legal property (example: data in private and non-private information). Our Italian scenario, CNR (a public body), noted (to the best of their knowledge) that the costs for public Cloud are still not eligible according to the Italian regulation.
- Financial, personnel (admin) and expertise in moving applications to the Cloud, financial issues, in regards to budget to pay for Cloud services.
- For those with Linux interests: flexibility, Software compatibility, and portability were noted.
- Deployment difficulty (is still not trivial)
- Intellectual property rights for research or industry discoveries
- Security when dealing with private data/data protection
- Availability of a complete and high reliable service, with all the required features
- Technical issues related to porting the applications so they can be used in a very User-friendly way
- Clear and complete documentation

Finally, below is a great comment addressing an aspect perhaps spoken far too little about:

“There is still a psychological jump needed by the scientific Users to see that Clouds are as reliable as local resources, and in many cases cheaper in cost (this last point is not clearly perceived since many maintenance costs are hidden to the scientist, who does not pay for power or technical staff).”
(Ignacio Blanquer- VENUS-C Scenario: Bioinformatics/Structural Analysis)

8a. What would be the best Cloud model for your application in the near future?
Rank each item from 1 to 5, five being the highest score.

a. Private Cloud (a private Cloud infrastructure by a university or company)
b. Community Cloud (a community Cloud by your application community)
c. Public Cloud (such as Azure)
d. A hybrid model of a private or community Cloud complemented by a public Cloud (a mixture of both services)

The rank(s) to receive highest score was:

a. Private Cloud - 33% of the respondents ranked 5 - the highest score
b. Community Cloud - 39% of the respondents ranked 4
c. Public Cloud – 28% of the respondents ranked 3 or 4 or 5. No one ranked 1 - the lowest score
d. Hybrid model – 39% of respondents for rank 5 – the highest score

Therefore, we could that the majority of the respondents believe that the community Cloud or hybrid model are the best models for their application in the future.

Figure 26 – Questionnaire Answer – Ranking of best Cloud model for your application in the near future

Of the elaborated answers, a common thread for those that prefer the hybrid Cloud is the benefit of accessing high computing power, but with restrictions and legal issues/dealing with sensitive data being better cared for.

- one respondent believes the hybrid Cloud is the most effective, as it can deal with low-scale experimentation and reuse local infrastructures.

In regards to private Cloud, noted was

- That resources would be provided somehow more easily managed in such a homogeneous environment
- It would be a good option in the early stages before moving to any other public Cloud.
- One respondent chose it as its first choice, if the costs were comparable, noting that there would be fewer issues to convince to potential collaborators/research projects if data/apps are "private".

In regards to the public Cloud,

- one respondent expressed the preference, as they need almost real time without long waits in queues for processing, and they need to share non-private data with others scientists.
A different respondent mentioned they keep experiment with the different models, and currently rank the public Cloud the lowest as they do not have means to account for it.

According to a respondent, the public Cloud could be the main provider of resources if financial issues and costs eligibility would be solved. Moving on to note that the private Cloud for them would be the only solution until the aforementioned issues are solved, or a new initiative similar to VENUS-C is funded.

9. Which 5 of the following Cloud Computing features do you find most attractive?
The top 5 most selected in order of highest percentage are:
Availability to scale up/down depending on the workload at 89%, Access to additional resources at 61%, Short time and cost for setting up new services at 56%, Access to on-demand (commercial) paid resources closer to deadlines tied with cost associativity at 44%, Ability to use software environments specific to my application at 39%.
(graph found within the document)

10. Would you recommend the Cloud service to a friend/colleague?
83% of the respondents said YES, while none said NO, leaving 17% at MAYBE.

Figure 27 – Questionnaire Answer – Whether the Cloud service would be recommended to a friend/colleague

Additional comments
Feel free to use space below to express any additional comments you may have. Any comment or suggestion is much appreciated.
(note: comments made are represented within the document)
ANNEX B - RESOURCE USAGE MODELS BY SCIENTIFIC DISCIPLINE

Below are a few charts representing the cumulative normalised CPU time of selected scientific discipline from June 2010 to May 2012.

Figure 28 – Cumulative Normalised CPU time of the Astrophysics discipline as part of EGI infrastructure (June 2011 to May 2012)

Figure 29 – Cumulative Normalised CPU time of the Earth Sciences discipline as part of EGI infrastructure (June 2011 to May 2012)
Figure 30 – Cumulative Normalised CPU time of the Computer Science and Mathematics discipline as part of EGI infrastructure (June 2011 to May 2012)

Figure 31 – Cumulative Normalised CPU time of the Fusion discipline as part of EGI infrastructure (June 2011 to May 2012)
Figure 32 – Cumulative Normalised CPU time of the High-Energy Physics discipline as part of EGI infrastructure (June 2011 to May 2012)

Figure 33 – Cumulative Normalised CPU time of the Computational Chemistry discipline as part of EGI infrastructure (June 2011 to May 2012)
• Cloud Computing has been forecast to play a big part in the future of ICT. Its potential is expected to be driven by the younger generations. In 2020, ICT will be worth 5 trillion dollars and at least 80% of industry will be driven by third-party platform technologies (browser-based, smart phones).
• What lies on top of the stack is important, not just the platform but ultimately the applications. Both research and industry adopters must ensure they carefully assess their investment with this mind. In the future, investment focus will need to shift towards applications and services across devices.
• The cloud is transforming the way people will work in IT, the skills and competences they will need. This evolving landscape means that applications are becoming increasingly device driven with an always-on mentality. People working in IT need to give up the idea of full control and focus on developing new skills, including contract negotiation.
• Building new IT skills comes hand in hand with fostering a cultural change in the way we deliver IT and services are vital for the wider uptake of Cloud Computing. It is important to grasp opportunities to re-train for the cloud ecosystem.
• Currently, we are seeing CIO playing the role of Chief Infrastructure Officer but over time this will become Chief Integrator Officer, and ultimately, Chief Innovation Officer
• The attention also needs to shift towards “easy to try, easy to buy”, scalable and supported interactions.
• The UK G-Cloud is demonstrating that operations and service deployment can be significantly accelerated thanks to the cloud. This is also another example for more focus on the application front, and not solely on infrastructure: we can hook silos together so people can build applications and create new services.

To unleash the full potential of Cloud Computing, Users should play a more prominent role in the standards process. To this end, standardization efforts need to become streamlined with more synergistic approaches to interoperability testing.
ANNEX D – GUIDELINES FOR RESEARCHERS AND THEIR INSTITUTIONS

The purpose of this annex is to provide pointers to newcomers to the cloud as well as for national resource providers and the role they can play in supporting the research community with recommended respective actions. This guide also takes into account that research computing in an institution is a complicated mix of funding, policy, process, academic needs and freedoms. Decisions may be informed by a number of factors, such as charging models, cross subsidies, carbon-reduction policies, the availability of academic discounts, all of which make it difficult to calculate all actual costs incurred.

Guide to end-Users from research

1. There is no ‘one size fits all’ – Assess your requirements

Cloud Computing is not a ‘magic wand’ or a universal solution that suits every type of IT environment or challenge. Cloud works best for applications and programming models that have been written specifically for the cloud. However, this is not yet common practice and many researchers may need to make adaptations. Individual researchers and research groups should make an assessment of their specific needs and application characteristics. Use cases and usage scenarios can serve as good examples in matching common requirements.

PRACTICAL STEPS:

✓ What do you want to move to the cloud?
✓ What requirements do you have?
✓ Are there any requirement forms, usage scenarios or guides available to facilitate your assessment?

2. IT skills vary – what technical support do you need?

Technical support for Cloud Computing is just as important as using a local cluster. Potential end-Users should carefully check what kind of technical support is needed. This support may be provided directly by the provider, by the institution or through web-based training.

PRACTICAL STEPS:

✓ Is support available at institutional level?
✓ Does the provider offer support?
✓ What on-line guidelines or courses are available?

3. Benefits differ – adjust your expectations accordingly

Potential benefits depend on the specific implementation of the cloud concept. Most researchers gain from the time it takes to complete a task using the cloud and new capabilities. Cost savings should not be a foregone conclusion. However, research benefits gained may outweigh higher costs than expected.

PRACTICAL STEPS

✓ Start with a free trial or competition if available
✓ Take a flexible approach to using the cloud and monitor tangible benefits and pricing as you go along
✓ Encourage your departments and institutions to collaborate on regular impact assessments, including cost-benefit assessments while facilitating your focus on core competences.
The nature and scale of Cloud Computing can accommodate tasks with large but transient loads. Advantages are most apparent when a large amount of computing power is required for a short time and/or when local facilities lack capacity or availability. Working with your department or institution means you can become more aware of the cost of resources consumed in the bigger picture, which, in turn, will encourage you to make a more efficient use of computing and computation infrastructure.

**Small businesses:** The steps are essentially the same as those describe in point 1 above. However, small businesses will be more concerned with pricing and keeping track of any special offers or price reductions by providers. In this respect, small businesses should also be aware that prices may vary geographically even from the same provider.

**TOP TIPS**
- Aim for the best value for money in terms of the benefits derived.
- Ensure the solution enables the focus to be on core research competences.
- Be realistic in your expectations.
- Ensure help is available if needed.

**Guide to institutions and funders**
Vice-chancellors and pro vice-chancellors, finance directors, directors of information services and systems, heads of academic and service departments, directors and managers of business change, including business analysts and enterprise architects need to respond effectively to changes in research funding and increasingly leverage opportunities to collaborate with businesses by ensuring flexible services.

Research councils and national funding agencies should treat Cloud Computing as a strategic planning problem and should examine the suitability of their policies and processes for a future in which cloud is an integrated part of the research computing ecosystem. Cloud Computing providers should be encouraged to participate in this process.

1) **Prepare for change**

JISC (UK) has provided a briefing on best practices guiding responses to funding changes with the aim of fostering efficiency and flexibility. This briefing is specifically targeted at decision makers in educational and research institutions. National funding agencies across Europe could draw on this guidance to reflect national priorities and specific circumstances, thus contributing to current best practices.

2) **Regularly assess the cost-benefits and impact of cloud usage in your institution**

Pricing and cost savings are not normally foremost in the minds of researchers when opting to migrate to the cloud. Further, requesting that they make such assessments means reducing the amount of time they can spend on core competences. Academic departments and institutions should encourage researchers to provide accurate, comprehensive and honest information on resources consumed. Decision makers (e.g. research managers and institutions) should work with researchers on performing their own price assessment of cloud services based on specific research tasks and circumstances. Specific research codes and tasks should be benchmarked to help determine how different cloud
offerings will perform against any available or proposed institutional infrastructure and therefore assess the relevant costs. Such assessments need to be revisited regularly as cloud offerings and prices change, when opportunities arise to make local facilities more efficient and as research computing requirements change. Comparisons should also take into account the potential benefits that Cloud Computing can offer in terms of the amount of computing that can be brought to bear on a task, and hence the time to complete a specific piece of research. Exposing researchers to the costs involved will raise awareness of their resource consumption in a wider setting and encourage them to make more efficient usage.

3) **Encourage investment in new skills/support staff**
Flexible and efficient services are synonymous with investing in the full set of “services” – technology, skills, and manpower. In order achieve these strategic goals, institutions need people who can bridge the gap between computing and researchers. New skills are needed for ‘moving up the stack’.

4) **Standardization for interoperability**
Education and research institutions have a key role to play in encouraging commitment to standards, interoperability and portability on the vendor side by fostering engagement with the standardisation process with the goal of accelerating consensus on relevant standards. It is particularly important that institutional decision makers encourage active engagement by end-Users moving forward.

5) **Support the creation of brokerage services at national level**
Brokerage services can play an important role by adopting a neutral position between providers and Users, contributing procurement frameworks and providing practical guidelines.

**Brokerage Services**
A body/interface that brokers with commercial suppliers on behalf of the research community would help address a number of current barriers to wider adoption. This type of body could provide arrangements to tackle security concerns; develop procurement frameworks to deal with complex legal and service level issues, including privacy, jurisdiction and liability questions under different legal systems and negotiate ‘on demand’ licensing of software used only temporarily in the cloud. Such a body should ensure a neutral position by working with both providers and end-User communities.

**Vast array of potential suppliers**
There is a bewildering array of options available for researchers thinking about using commercially provided Cloud Computing. Amazon alone has eleven instance types currently listed with prices varying according to geography. Most individual researchers do not have the time nor inclination to properly analyse and compare the different options – and price may not necessarily be the most important factor in deciding which provider to use.

Brokerage services should provide up-to-date information on current offerings and related costs to facilitate both individual researchers, their departments or institutions in selecting the best provider for the task at hand.