e-FISCAL

Financial Study for Sustainable Computing e-Infrastructures

Deliverable Title: D. 2.3 Computing e-Infrastructure cost estimation and analysis – Pricing and Business models

Partner Responsible: AUEB-RC

Work Package 2: Computing e-Infrastructure cost calculation

Submission Due Date: 31.1.2013

Actual Submission Date: 1.3.2013

Amendment Date: 3.7.2013

Distribution: Public

Nature: Report

Abstract: This document provides an overview of the European HTC and HPC costs calculated with the e-FISCAL hybrid methodology, while comparing these costs with commercial cloud offerings. It also presents the developments of business and pricing models for the use of commercial resources. Finally it provides an evaluation of the related estimations and concludes with recommendations for various stakeholders.

Document Information Summary	
Deliverable number:	*D2.3*
Deliverable title:	*Computing e-Infrastructure cost estimation and analysis – Pricing and Business models*
Editor	*Sandra Cohen*
Contributing Authors:	*Sandra Cohen, Fotis Karagiannis, Costas Courcoubetis AUEB-RC, Kashif Iqbal ICHEC, Sergio Andreozzi EGI.eu, Matti Heikkurinen, ETL*
Reviewer(s):	*Sy Holsinger EGI.eu, Matti Heikkurinen, ETL*
Work package no.:	*WP2*
Work package title:	Computing e-Infrastructure cost calculation
Work package leader:	*Sandra Cohen*
Work package participants:	*Sergio Andreozzi, Panagiotis Antonopoulos, Owen Appleton, Costas Courcoubetis, Matti Heikkurinen, Sy Holsinger, Kashif Iqbal, Fotis Karagiannis, Steven Newhouse, George Thanos,*
Distribution:	*Public*
Nature:	*Report*
Version/Revision:	*v.am1*
Draft/Final	*Final*
Keywords:	*Cloud offering, pricing models, business models,*

DISCLAIMER

Grant Agreement Number: 283449
Start Date: 1 August 2011
Duration: 18Months

Log Table

Issue	Date	Description	Author/Partner
v1.0	14/11/2012	Skeleton	Sandra Cohen /AUEB
V2.0	23/1/2013	Updated skeleton	Sandra Cohen/AUEB
V2.1	25/1/2013	Updated skeleton	Sandra Cohen/AUEB
V3.0	7/2/2013	Updated skeleton	Sandra Cohen/AUEB
V3.1	15/2/2013	First draft – incomplete	Sandra Cohen/ AUEB
v.4.0	19/2/2013	First draft – complete	Sandra Cohen/AUEB Fotis Karayiannis/AUEB *Matti Heikkurinen/ETL* *Kashif Iqbal /ICHEC* *Sergio Andreozzi/EGI.eu* Sy Holsinger/EGI.eu
v.4.1	19/2/2013	Incorporation of contributions	Sandra Cohen/AUEB
v.4.2	20/2/2013	Several updates	Sergio Andreozzi/EGI.eu *Kashif Iqbal /ICHEC* Sandra Cohen/AUEB
v.4.3	20/2/2013	Version sent to the AB	Fotis Karagiannis/AUEB
v.4.4	25/2/2013	Editing changes based on various comments	Sandra Cohen/AUEB
v.4.5	27/2/2013	Document review	Sy Holsinger/EGI.eu
v.4.6	27/2/2013	Document review	*Matti Heikkurinen/ETL*
v.4.7	27/2/2013	Integration of reviews	Sandra Cohen/AUEB
v.4.8	27/2/2013	Edition and review taking into account the Advisory Board comments for PMB approval	Fotis Karagiannis/AUEB
v4.9	28/2/2013	Integration of comments	Sandra Cohen/AUEB
v5.0	28/2/2013	Comments	Sergio Andreozzi/EGI.eu
v5.1	28/2/2013	Version for PMB approval	Fotis Karagiannis/AUEB
v.5.2	1/3/2013	Integration of PMB comments	Sandra Cohen- Fotis Karagiannis/AUEB
v.final	1/3/2013	Final version submitted to the EC	Fotis Karagiannis/AUEB
v.am1	3/7/2013	Amended version based on reviewers comments	Sergio Andreozzi/EGI.eu, Sandra Cohen- Fotis Karagiannis /AUEB, Matti Heikkurinen/ETL

Document Approval

Approval by	Date
Project Management Board	3/7/2013

EXECUTIVE SUMMARY

Over the last ten years the European countries and the EC have made significant investments in e-Infrastructures for scientific computing, notably High Throughput Computing (HTC) and High Performance Computing (HPC) services. Their contributions have made it possible for the European research to maintain most prominent role in solving key global challenges. The approach has been validated by a growing number of research initiatives and success stories such as the recent discovery of the Higgs boson. Sustainability of such services is essential, as the research supported by them is also more and more crucial for European competitiveness – however, sustainability can only be planned if the costs are known. The emerging commercial offerings (Cloud-based HTC and HPC solutions) pose additional opportunities and challenges for sustainability. It is thus important to understand the cost of the dedicated computing related e-Infrastructures, as primarily expressed by EGI and PRACE initiatives, both in their current state and in their evolution towards the Horizon 2020 framework. The calculation of such costs is not trivial; keeping detailed accounting data for the evolving mix of capital (hardware) and operational (personnel, energy) expenditures is only one of the challenges. This is because, each of the compared infrastructures among the HPC, HTC and Cloud offerings have their distinct characteristics and use-case applications. For example, high-end HPC systems tailored for capability computing deploy state of the art hardware and network as compared to HTC infrastructure.

The goal of the e-FISCAL project has therefore been to analyse the costs of the current European dedicated HTC and HPC computing e-Infrastructures for research and compare the service components provided with equivalent commercial leased or on-demand offerings. With regard to the comparison, it is crucial to keep in mind that the service palette provided by the computing e-Infrastructures – which includes advanced user support, engagement with user groups, cross-organisational authentication and authorisation mechanisms – extends beyond the basic Cloud computing service used in the comparison. Furthermore, the results should not be used to evaluate moving from in-house to commercial public clouds; this is a different exercise and is a next step requiring a different approach based on avoidable costs, while this is also related to the qualitative value of e-Infrastructure that cannot be easily measured. Still, this cost comparison allows a baseline quantitative analysis of the cost ratios that is more easily repeatable than comparison addressing broader set of services and their qualitative aspects.

To perform the cost analysis, the project developed a hybrid costing methodology that builds on the two main methodologies used for cost assessment in e-Infrastructures: Total Cost of Ownership (TCO) and Full Cost Accounting (FCA) that have been used in most of the case studies identified in the state of the art analysis. However, neither of the methodologies was sufficient for reaching the e-FISCAL goals. While TCO is a useful concept in assessing the cost of a specific project over its useful life, it sets high demands for the analytical model and data used to make predictions. In a heterogeneous distributed e-Infrastructure it is very hard to come up with acceptably precise results by using a methodology that is almost completely forward-looking. On the other hand, FCA methodology relies on actual cost accounting data information that is dependent on the level of cost accounting systems sophistication by additionally inducing a backward looking stance, and gathering this type of information in Europe-wide infrastructure with sufficient accuracy and coverage is impossible.

The e-FISCAL model is a hybrid model that approximates the costs of maintaining services at their current level in the short to medium term. It does that without the need to identify funding sources or the exact time when the actual infrastructure investments have been made. Therefore, rather than relying on the detailed financial data or projections to the future, it uses high level information about computing and storage hardware costs (including interconnection costs), auxiliary equipment costs (i.e. cooling, uninterruptable power supply devices - UPSs, power generators), software costs, personnel costs, and site operating costs. The e-FISCAL methodology is completed in two phases; firstly a simulation of the physical infrastructure is sketched and secondly the annualised cost of the simulated physical infrastructure and the operating cost of the physical infrastructure are added together. All the data necessary to feed the model (e.g. e-Infrastructure acquisition costs, personnel costs, electricity cost, depreciation rates) is retrieved from a properly developed questionnaire (the e-FISCAL survey instrument).

In the study, the e-FISCAL methodology has been applied to a sample of HTC and HPC centres in Europe making this the first initiative to gather and analyse the costs from a representative number of countries in a comprehensive and systematic way. As is evident by the state of the art review, the majority of costing studies either concentrate on a site or make a multiple site albeit in the same country comparison. The analysis revealed that the cost per core hour for 2011 lies in the range of €0.03 core hour to €0.07 core hour[1].

HOW TO READ THE RESULTS

The cost analysis in this project has been undertaken under three alternative approaches – basic case (costs for whole e-Infrastructure), basic case split (cost calculated for HTC and HPC services separate) and case by case analysis (per computing centre approach) – with results calculated for both yearly cost per core and cost per core hour. To preserve anonymity of the answers, results are discussed consistently using average and median values. As there is a significant difference between median and average values, it is useful to discuss this difference and help managers to interpret the results. Average is the mean value taking all sites equally into account regardless of how far from the "norm" they fall. In contrast, the median value has the inherent property of excluding the "extreme" values (high or low), and thus represents the "most likely value for a typical case" that is more appropriate for guidance and decision making. Median is thus a more relevant reference point for both computing centre managers and policy makers.

Average values are always higher than their median counterparts. This is an indication that there are computing centres that are less cost- efficient when measured purely in terms of cost per core hour, and that there is space for improvement in order a site to move towards the cost that corresponds to the median value. Sites that are close to (or below) the median value should be considered as cost efficient compared to the group of academic installations in our sample. Sites above the average value should try to analyse the issues that affect their cost efficiency, and try to separate activities providing genuine added-value (that may not be captured by the cost per core hour metric) from the areas where efficiency improvements would be possible without adverse consequences.

[1] Note that the above values takes into account all the personnel working directly or indirectly in the centre, namely for administration, operation, middleware and application development, user support, training, dissemination, as well as managers and policy makers

Some other underlying issues revealed by the analysis were:

- Hardware depreciation rates are in several cases well above the typical well-cited three–year period. The average depreciation period corresponds to 5 years,
- The Power Usage Effectiveness (PUE) rates are around 1.5 (median value) indicating rather efficient energy use, and
- Approximately 2 FTEs are required for each 1,000 cores (median value).

The analysis also provided evidence of a decreasing trend in costs from year 2010 to 2011.

After assessing the per core hour costs, they are compared with the prices charged by commercial cloud providers (such as Amazon EC2). The comparison is done in two variants: first without taking into consideration performance differences between in-house service provision and commercial cloud services, and second by adjusting for the differences in performance based on the benchmarking results (performance-adjusted cost comparisons). The small-scale benchmarking exercise included in the project was used to establish range of performance degradation percentages that correspond to the most relevant in-house infrastructure installations and their commercial service counterparts.

By concentrating only on financial considerations, our findings provide strong evidence that **cloud prices are not necessarily** and unanimously **lower than** the costs calculated by e-FISCAL for **in-house service** provision. This holds true even without taking into account the differences in the actual service portfolios discussed above. Parameters like utilization rates, infrastructure size, length of time committed to use cloud services (for example by using "reserved instances" offered by some of the Cloud vendors), efficiency in manpower utilization and service performance are important factors affecting the basic cost relation between cloud and e-Infrastructures. Thus cost is not synonymous to value; on the contrary **cost is only one of the parameters that should be considered in the value creation process**.

Main achievements and findings

During the final part of the e-FISCAL project, the consortium consolidated its findings focusing on the cost estimations of the European computing e-Infrastructures based on survey answers and other relevant data (market and literature). Estimates of basic metrics, such as the cost per core hour, have been updated based on a slightly extended sample (basic case[2]) as well as based on a new methodology (the case-by-case analysis[3]), which narrowed down the result ranges. Furthermore, the project calculated separate values for HPC and HTC (basic-case split[4]) based on the corresponding sub-samples (using the respondents own feedback regarding their categorisation of HPC or HTC (or both), we call this analysis basic-case split). We used the number of EGI cores in order to approximate the total yearly cost of the dedicated European HTC e-Infrastructure and the number of PRACE Tier-0 centres' cores together with other national Tier-1 centres' cores in order to approximate costs of the dedicated European HPC e-Infrastructure. The number of cores is based on figures reported in public deliverables or project websites, and analysis is based on the values 2011 (the latest year in the survey scope)[5]. The benchmarking results with the NPB and HEPSEPC06 benchmarks have been integrated in the costing exercise by introducing the performance degradations through weight factors, and updated (and fairer) comparisons have been made with commercial cloud providers, adding OVH.com besides Amazon. Sensitivity analyses have been performed showing interesting results, in particular illustrating the crucial role of the utilisation rate of the in-house infrastructure and comparing the results with the different instances from Amazon. Moreover, the project extended to other areas such as business models for the use of commercial resources serving the research community and possible pricing schemes for leasing resources. The main results achieved during the second reporting period can be summarised as follows:

- ***Update of the State of the art***: The project identified and *reviewed* a total of *50 related publications and articles (another 11 since the first period)* with emphasis on business and pricing models. The state of the art review is published on the project website and has proven to be an important tool for engaging with the researchers and broader audience. The link to the e-FISCAL state of the art repository from Wikipedia page on Cloud Computing[6] remains an important source of visitors to the project website.
- ***Updated cost estimations for European HTC/HPC e-Infrastructure***:
  The **sample** for the e-FISCAL financial study has been *extended with an additional 4 responses including 2 new countries (Italy and the UK*) bringing the total sample up to 28 *responses from 16 countries*. However, some of the largest HPC centres (such as PRACE Tier 0s) are not included in the survey sample due to confidentiality reasons (such as non-disclosure agreements between the vendors and these centres). The *e-FISCAL values have been calculated as follows:*
  - **Basic case**: The updated median values (which reduce the impact of the so-called "outliers" that are either very high or low compared to typical values) are around €0.03 per core hour in 2011, while averages are around €0.07 per core hour in 2011. The detailed values are presented in the section 4 of this document.

[2] Results presented on Table 2 on page 43
[3] Results on Table 6 on page 45
[4] Results on Table 4 and Table 5 (page 45)
[5] However the performance of PRACE Tier-0 systems has not been taken into account in the study
[6] http://en.wikipedia.org/wiki/Cloud_computing

- **Basic case split (to HPC and HTC)**: The corresponding values for the HPC centres (self-identified as providing HPC services) are slightly lower than €0.03 per core hour for the median and €0.05 per core hour for the average. For the HTC centres (also self-identified) the median is between €0.03 per core hour and €0.04 per core hour, while the average is close to €0.08 per core hour.
- **Case-by-case analysis**: In this case, the e-FISCAL costing model is run on each of the centres based on their reported input values. In case of missing values the average values from the whole sample were used. Based on this approach, the median cost per core hour is roughly € 0.06, while the median cost for HPC core hour is slightly below €0.04 per core hour and for HTC between €0.06 and €0.07.

All of these findings are in-line with the published results (that have been included in state of the art repository). Some key observations are listed below:

- The breakdown between CAPEX and OPEX in 2011 in our calculations is around 30%-70% respectively (median) to 26%-74% average. 51% total costs (median) is dedicated to personnel costs.
- The utilisation rate used to calculate the average and median cost per core hour for the above results for 2011 is 65% and 75% respectively.
- Other findings include the slow depreciation rates typically used for the hardware (average 5 years), the PUE values of around 1.5 median value (lower than typical rates quoted for the industry average) and the percentage of electricity cost (around 16-17% median value of all costs).

- ***Approximated annual total cost of the in-house HPC and HTC e-Infrastructure in Europe*** *lies between 175 and 295 M€.* In order to approximate the cost of HPC and HTC centres we use the number of cores available in 2011 and we multiply with the average and median cost per core per year under option 2 (basic case split) and option 3 (case by case analysis) summarised above. The numbers of cores in the dedicated HTC e-Infrastructure (EGI) is roughly 300.000, while the dedicated HPC e-Infrastructure (PRACE Tier 0's and other national Tier 1's) consists of approximately 550.000 cores (both in 2011). Hence the yearly cost of the dedicated HTC e-Infrastructure lies between 70-126 Million Euros, and between 103-170M€ for the dedicated HPC e-Infrastructure that includes Tier 0 and Tier 1 HPC systems[7]. The latter estimate is based on the e-FISCAL sample which is populated by Tier 1's and not Tier 0's.
- ***Comparison with commercial on-demand prices*** was informed taking into account the results of the benchmarking efforts of the project on the latest Amazon EC2 pricing information[8]. The comparison focuses on pure computing costs and omits some of the other costs categories: network and storage costs are not included (they depend on usage patterns of the application), personnel (such as application developers and administrators) for configuring and operating EC2 instances and for the adaptation of the application code

[7] The difference in the performance between HPC systems has not been taken into account and findings are again based on e-FISCAL models and data

[8] Amazon dropped EC2 prices an additional 2 times, with latest change on 1st February 2013. However, the pricing on the 31st January was retained as the basis for comparison, as this was the last day of the project and already gave considerable advantage since the prices were compared with costs based on 2011 information.

that would be needed if researchers migrate to Amazon. Comparing in-house costs with EC2 prices[9] brings in an inherent uncertainty (not to mention costs from 2011 with prices in 2012/2013).

The comparison is not straightforward as there are several types of instances such as reserved (for 1 and 3 years) and on-demand, or low-end and high-end, which have significant differences among them. Furthermore, the utilisation of the reserved instances needs to be taken into account, since it might not be possible to absorb all of the Amazon reserved capacity in all occasions. A sensitivity analysis has also been performed having as a variable the utilisation rate of the e-FISCAL in-house infrastructure (supposing all the rest remain constant) and comparing them with the Amazon ones. Some basic conclusions (once again taking into account the benchmarking results) are:

- EC2 "on-demand" instances are always more expensive than e-FISCAL in-house findings (even "standard" Amazon instances like L/XL (Large/Extra Large).
- EC2 "Reserved" instances are much more competitive than "on-demand" ones and are comparable to e-FISCAL (and in fact towards the e-FISCAL low end, i.e. the € 0.03 per core hour).
- However EC2 "cluster compute" (HPC in the cloud instances) are 1.5 to 2 times more expensive than the e-FISCAL costs per core hour

Regarding the sensitivity analysis on the utilisation the main conclusions are (again there are variation among the different EC2 instances) summarised as follows. The in-house utilisation rates above which it is more cost-efficient to keep the infrastructure in-house are:

- 30-60% (median-average) compared to EC2 "standard on-demand" instances.
- 15-25% (median-average) compared to "on-demand cluster compute".
- 40-95% compared to "standard reserved ones" and in one case (the 3 year standard reserved ones compared to e-FISCAL average) Amazon is always more cost-efficient from in-house even with a 100% utilization for the in-house.
- 20-40% for 1 year reserved "cluster compute" ones.

This can be further summarized that the cluster compute instances are not yet comparable and it is more cost-efficient to keep them in-house (even with a low utilisation 15-40%). While for the standard EC2 ones, on-demand are again less cost-efficient (but not at the level of cluster compute ones), while 1-3 year reserved ones are comparable to e-FISCAL and sometimes can be always cheaper (even compared to 100% of in-house utilisation).

- ***Review of business models*** highlighted the trend to combine the usage of public-funded federated infrastructures with commercial cloud providers to expand the offered capacity, to enable access to different resources or service levels or to enable efficient access to research data for commercial exploitation. Furthermore, pricing models characteristics from cloud services show the way to define pricing schemes for leasing in-house e-Infrastructure services (if legally permitted). The identified pricing models and classification have been derived by analysing the real-world offering from the major commercial cloud providers. Usually, all providers offer at least usage-based and subscription-based models, while only one provider digs into the market-based model. Strategy-based models may appear for certain customer domains (e.g. education, not for profit). Some key conclusions are the following:

[9] Pricing is influenced by the profit margins, which in some cases can also be negative.

  - High utilisation is key to maintain economic efficiency and thus strongly related to the business model that will be selected in a hybrid in-house and leased infrastructure.
  - A broker role is essential to facilitate demand meeting the right suppliers and to increase utilisation
  - Need to evolve funding streams and new pricing models.
- **Green IT**: Green IT is a very wide topic, with multitude of approaches. However, the PUE ratio is at the moment the most universally known single metric, and in addition to indicating technical efficiency of operations it may also indicate the degree of attention given to the broader Green IT issues.
- ***e-FISCAL workshops***: The above results have been presented and discussed initially at the e-FISCAL workshop in **Prague** and then the final e-FISCAL workshop in **Amsterdam**[10], where very positive feedback has been received from financial experts on the results.
- **e-FISCAL tools and sustainability**: e-FISCAL prepared not only the promised downloadable spreadsheet tool, but also an on-line version of the tool (only marginally funded by e-FISCAL) that will help any computer centre manager to self-assess its centres' costs, **new or existing**, and compare them with the e-FISCAL average and median values. Both are available under www.efiscal.eu/tools

  Regarding the e-FISCAL outputs sustainability:
  - The tools will remain available on the website under free type licenses.
  - State of the art repository will also remain available and updated on a best effort basis.
  - LinkedIn group on ICT cost assessment[11] continues discussions.
  - Cost collection/estimation to continue in part through the EGI compendium (go.egi.eu/EGI-Compendium-2011).
  - Consortium members can provide consultancy upon request.
- ***Regarding the e-FISCAL potential impact and use,*** e-FISCAL will continue to contribute to the continuity and sustainability discussions for e-Infrastructures as was done with EGI and its "Evolving EGI" workshop, with which the final e-FISCAL workshop was combined. The identified costs calculations and business models contribute in this direction. Better understanding of the financial realities related to computational e-Infrastructure services allow more efficient strategic shaping of incentives on national and European scales to drive innovation related to e-Infrastructures and the research activities supported by it. e-FISCAL claims that it laid the ground for better resource utilisation, provided the related financial backing for growing new markets and expanding investments.
- **The sustainability of the e-FISCAL effort** depends to a high degree on the number of individuals and organisations that adopt it in the future activities. The success in this area is based on the continued use of the approach in the EGI Compendium effort as well as choosing a licensing scheme that maximises the reuse and visibility of the results.
- **Satisfaction surveys** were also sent to the teams of the computing centres that answered and up to the time of the writing all answers received (19) indicate the e-FISCAL effort as useful (11) or very useful (8). The scale was from (1) major waste of effort to (5) very useful. Some of the comments received are highlighted below:

[10] Reports from the e-FISCAL workshops are available: http://www.efiscal.eu/events

[11] LinkedIn group on ICT cost assessment: linkd.in/VqEth0

- The e-FISCAL approach to analyse trends and compare different market segments makes it a useful tool in understanding strengths and weaknesses of our own organisation and its role in the marketplace.
- We used the e-FISCAL work to compare with our local systems (and they were very comparable)
- I was really glad to see that a couple of beta computation tools have been included in project's website http://www.efiscal.eu/tools.
- The online tool is great.
- Carrying out personal interviews would probably make information gathering more efficient and complete.
- It would be useful to have structured method for HPC centres to follow so they can calculate their operational costs. This may be a set of guidelines which HPC centres have to go through (check)
- The outputs were useful. However it is very difficult to make a step forward towards sustainability for the distributed grid infrastructures due to various reasons, e.g. no clear funding path at national level, no clear sustainability path at European level (for both EGI and the "evolution" of EMI), no concrete actions to make the current distributed infrastructures available (and appealing) to new user communities (in particular the ESFRI projects).
- I would like to highlight the importance of maintaining this information updated in the coming years, in order to follow the evolution of "this market". The methodology developed under e-FISCAL, if continued in time, would provide relevant information to guide investments in e-infrastructures. Supercomputing centres need clear, reliable, and updated information to optimise investments in infrastructures like the one that e-FISCAL attempted to provide.

TABLE OF CONTENTS

TABLES

FIGURES

GRAPHS

1. Introduction

1.1 Scope of the document

This document constitutes the final deliverable of the e-FISCAL project. As such, in addition to the actual results, it encapsulates all the activities, approaches and tools relevant to e-Infrastructure cost estimation and analysis. The basic goal of the e-FISCAL project is to analyse the costs of the current European dedicated HTC and HPC computing e-Infrastructures for research and compare them with equivalent commercial leased or on-demand offerings. In order to achieve this goal, e-FISCAL has developed a costing methodology that fits its purpose and has gathered information through a survey instrument (questionnaire) from 28 respondents that correspond to 16 countries. It should be noted that the results should not be used to evaluate moving from in-house to commercial public clouds; this is a different exercise and is a next step requiring a different approach based on avoidable costs, while this is also related to the qualitative value of e-Infrastructure that cannot be easily measured.
The research questions that are addressed through this project are summarized in the following five:

- Research question 1: To analyse the cost structure of HPC and HTC centres
- Research question 2: To approximate the cost of HPC and HTC in Europe
- Research question 3: To compare the cost of HPC and HTC centres with commercial Cloud Offerings
- Research question 4: To propose suitable business models for e-Infrastructures
- Research question 5: To analyse Green IT initiatives in HPC and HTC centres

However, not all research questions have equal importance in the study. The first three questions correspond to the main focus of the project while the final two issues are covered less extensively as they are addressed in-depth by other European projects and initiatives as described in section 4.

In the document, we present the main findings evident in the state of the art in relation to the research questions raised. We then describe the methodology followed in order to calculate the costs of dedicated HTC and HPC computing e-Infrastructures as well as the results of the application of the methodology and the costs derived. We present the cost per core hour as well as the cost structure of HPC and HTC sites for 2011 (e.g. cost break down CAPEX/OPEX) under alternative calculation approximations. Moreover, we compare the cost of in-house computing provision with the prices of cloud commercial offering both before and after adjusting for difference in performance between in-house computing and cloud services. The difference in performance has been assessed through a benchmarking exercise executed within the project that is analytically discussed in this report. We also discuss business models and pricing policies. Finally, we use PUE numbers as the best proxy for the broader Green IT initiatives adopted by HPC and HTC centres and we therefore gather such information from our sample sites. The document also concludes with recommendations and proposals that stem from the findings and experiences gained through the process of project materialization.

1.2 Target Audiences

The document is intended as an internal and external reference. Internally its main purpose is to collect and transfer knowledge to various follow-up activities, while at the same time also being of interest to a number of external stakeholders such as:

- e-Infrastructure computing providers and HPC/HTC centre managers, providing them with a better understanding of the cost structures and actual costs of in-house vs. leased infrastructures to assist in planning and determining their centres' sustainability.

- Research communities that can benefit from cloud computing allowing them to compare their costs against in-house infrastructures.
- e-Infrastructure policy makers and funding agencies, who recognise the cost parameter as a key parameter in the sustainability of computing e-Infrastructures.

1.3 Structure of the document

The document has six (6) sections and eight (8) Appendices.

- **Introduction** is found in Section 1.
- In Section 2 we discuss the **state of the art**. State of the art review covers costing models for e-Infrastructures, comparisons between commercial clouds and in–house computing in several fronts, business models for e-Infrastructures and Green IT issues.
- Section 3 is devoted to the **methodology** followed in the project. It thoroughly presents the costing model developed for the scope of this project, the research setting, the questionnaire used to gather cost information from respondents, the questionnaire dissemination and follow up as well as the characteristics of the benchmarking exercise.
- Section 4 presents the **results** of our analysis (i.e. cost per core hour, cost structures, comparing e-FISCAL costs with commercial cloud providers' prices, business models and pricing models identified, etc.).
- Section 5 provides a brief **discussion** of the main findings of our project and concludes with **recommendations** to the main stakeholders.
- Section 6 hosts the list of **references** of resources used in the report.

The appendices provide additional information and data. More specifically:

- In Appendix 1 the final e-FISCAL questionnaire administrated through the SurveyMonkey tool is found as a pdf.
- In Appendix 2 a lighter version of the e-FISCAL questionnaire developed as a spreadsheet along with the web-based tool can be found.
- In Appendix 3 a step-by-step guide on how the e-FISCAL cost model works is illustrated.
- In Appendix4 the sample characteristics as well as descriptive statistics of the input received by respondents are disclosed in a way however that anonymity is preserved.
- In Appendix 5 we present the results of univariate Ordinary Least Square (OLS) regressions where the cost per core, the cost per core hour and the number of FTEs/1,000 cores is regressed over the number of 1,000 cores.
- In Appendix 6 we compare the prices of Amazon for Windows instances with e-FISCAL costs.
- In Appendix 7 we compare the technical information regarding the benchmarking exercise.

1.4 Terms and definitions

Capital Expenditures (CAPEX)	Capital Expenditures are costs incurred in order to create future benefits. This cost is accounted for during the periods the assets are economically used through depreciation.
Cloud computing	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.
Central Processing Unit (CPU)	A central processing unit (CPU), also referred to as a central processor unit, is the hardware within a computer that carries out the instructions of a computer program by performing the basic arithmetical, logical, and input/output operations of the system.
EGI – European Grid Infrastructure	A federation of shared computing, storage and data resources from national and intergovernmental resource providers that delivers sustainable, integrated and secure distributed computing services to European researchers and their international partners.
EGI.eu [12]	A non-profit organisation based in Amsterdam established to coordinate and manage the infrastructure (EGI) on behalf of its participants: National Grid Initiatives (NGIs) and European Intergovernmental Research Organisations (EIROs).
Full Cost Accounting (FCA)	Full Cost Accounting is a typical costing methodology within which detailed actual data (in line item format) is attributed and allocated on the basis of specific algorithms and various costing procedures to services/products in order to come up with the "cost per unit" of the object being analysed.
HPC- High Performance Computing	HPC is a computing paradigm that focuses on the efficient execution of compute intensive, tightly-coupled tasks. Given the high parallel communication requirements, the tasks are typically executed on low latency interconnects which makes it possible to share data very rapidly between a large numbers of processors working on the same problem. HPC systems are delivered through low latency clusters and supercomputers and are typically optimised to maximise the number of operations per seconds. The typical metrics are FLOPS, tasks/s, I/O rates.
HTC - High Throughput Computing	HTC is a computing paradigm that focuses on the efficient execution of a large number of loosely-coupled tasks. Given the minimal parallel communication requirements, the tasks can be executed on clusters or physically distributed resources using grid technologies. HTC systems are typically optimised to maximise the throughput over a long period of time and a typical metric is jobs per month or year.
Hyper-threading	Hyper-threading (officially Hyper-Threading Technology or HT Technology, abbreviated HTT or HT) is Intel's term for its simultaneous multithreading

[12] http://www.egi.eu/

	implementation first appearing in February 2002 on its Xeon server processors and in November 2002 on its Pentium 4 desktop CPUs. Later, Intel included this technology in Itanium, Atom, and Core 'i' Series CPUs, among others. Intel's proprietary HT Technology is used to improve parallelization of computations (doing multiple tasks at once) performed on PC microprocessors. For each processor core that is physically present (physical core), the operating system addresses two virtual or logical cores, and shares the workload between them when possible. The main function of hyper-threading is to decrease the number of dependent instructions on the pipeline.
IaaS- Infrastructure as a Service	One of cloud computing models; in this most basic cloud service model, cloud providers offer physical or more often as virtual computing cores (virtual machines), storage, network access and other infrastructure resources. IaaS providers supply these resources on demand from their large pools installed in data centres.
Logical processor or logical CPU[13] or logical core or CPU core	A processor (CPU) that handles one thread of execution (instruction stream). A logical processor can be a (physical) core or a hyper-thread. There can be one or more logical processors per (physical) core (more than one if hyper-threading is enabled) and one or more cores per processor socket.
NGI - National Grid Initiatives	NGIs are the entities responsible of procuring and operating the national grid infrastructure (in terms of computers and storage devices) and corresponding services to the research and academic communities.
Operating Expenses (OPEX)	Operating expenses refer to expenses incurred in the ordinary course of business, such as salaries, administration and selling expenses, energy expenses, overhead, etc. These expenses are considered costs when they incur.
PRACE (Partnership for Advance Computing in Europe)	PRACE is a unique persistent pan-European Research Infrastructure for HPC implementing 15 petaflop supercomputing systems in Europe. PRACE manages extreme computing power and a selected set of highly specialized services.
Total Cost of Ownership	TCO is a methodology used in order to assess the cost of a specific project over its useful life. All expected costs over the project's lifetime are summed up and then divided by the anticipated useful life of the project to come up with a yearly cost. TCO is commonly used as a basis for calculating the costs of e-Infrastructures

[13] http://msdn.microsoft.com/en-us/library/dd722831%28v=bts.10%29.aspx

2. State of the Art

The state of the art review was initiated as the first step within the project and continued throughout the lifetime of the project. It included review of publications in several domains, such as costing models for e-Infrastructures, cloud computing, business models and Green IT. However, costing models for e-Infrastructures and cloud computing did have the most direct link with the core research question, hence publications in those domains have been reviewed most extensively.

The state of the art review was conducted through usual desk research methods and by attempting to contact authors of the papers found in order to see if they had additional suggestions. This engagement with the authors produced additional references. The desk research was aimed at finding relevant project reports, conference papers, academic papers (accessed through university subscriptions), Internet articles, professional reports, etc. The relevant documents reviewed are published in the state of the art repository (http://www.efiscal.eu/state-of-the-art). At the end of the project the e-FISCAL repository counts 50 references.

2.1 Costing models

The review of the state of the art revealed several categories of studies relevant to our project under the lens of costing models which are explained below.

2.1.1 Cost categories

A first category of papers deals with the typical cost categories that should be included when analysing the cost of computing e-Infrastructures. The review of the state-of-art-literature revealed that there is a rather common break down of HTC/HPC costs into the following categories: 1) computing and storage hardware costs including interconnection costs, 2) auxiliary equipment costs (cooling, UPS, power generator), 3) software costs, 4) personnel costs, 5) site operating costs (including electricity and premises costs) and 6) connection costs. This cost breakdown is stated in Opitz et al (2008) and has been repetitively adopted in several other studies. Kashef and Altmann (2012) review several papers and propose a similar categorization as that of Opitz et al. (2008). These cost categories fall within two main cost distinctions: Capital Expenditures (CAPEX) and operating expenses (OPEX). Capital Expenditures are incurred in order to create future benefits e.g. assets acquired have a useful life beyond one year. The cost is accounted for during the periods the assets are economically used through depreciation. Operating expenses refer to expenses incurred in the ordinary course of business, such as salaries, administration and selling expenses, energy expenses, overhead, etc. These expenses are considered costs when they incur. 360 Intersect in the HPC Budget Allocation Map -Willard et al (2013)- have also very similar categorization although naming may vary slightly.

2.1.2 Costing methodologies

A second category of papers dealt with the costing methodology to be applied in order to calculate yearly costs. There are several cost accounting methodologies in order to perform a costing exercise e.g. activity based costing, life cycle costing[14], total cost of ownership (TCO)[15], full cost accounting (FCA)[16], etc. Generally speaking, selection

[14] Life Cycle costing is a tool to calculate the economic costs caused by a product or a service during its entire life cycle, from purchase of raw material and components, cost of production and investments to usage, maintenance and waste management. EU had commissioned the development of elaborated guidance on the application of life cycle costing and on

of costing methodologies is heavily reliant on the industry on which the costing exercise is performed. For example, manufacturing companies can choose between a plethora of existing techniques such as process costing, job-order costing, standard costing, marginal costing, target costing, activity based costing[17], etc. In electronic communications, the accounting method of LRAIC (long run average incremental cost[18]) methodology that is rather not encountered in other business setting is highly demanded from regulators for regulated wholesale services costing. Service companies with high percentage of personnel cost, are more keen on adopting Activity Based Costing (Palaiologk et al., 2012[19]). Also, the selection of the costing methodology depends on whether the performer of the costing exercise lies in or out of the company. Costing information is sensitive and companies as well as institutions are rather reluctant to disclose it in an analytical form. Therefore, external analysts can only rely on basic information and then apply a set of hypotheses and assumptions in order to conclude on a cost output. Finally, a costing exercise could follow either a bottom–up or a top-down procedure. The bottom-up (BU) approach develops the cost model on the basis of the expected demand for the service provision of the infrastructure and sets the infrastructure design and estimates the related costs on the basis of an engineering model. In a top-down (TD) model, the starting source of information is the cost actually incurred by the infrastructure derived from the institutions accounts.

The most relevant costing methodologies in computing e-Infrastructures are TCO and FCA. Therefore we concentrate on these two in the following discussion.

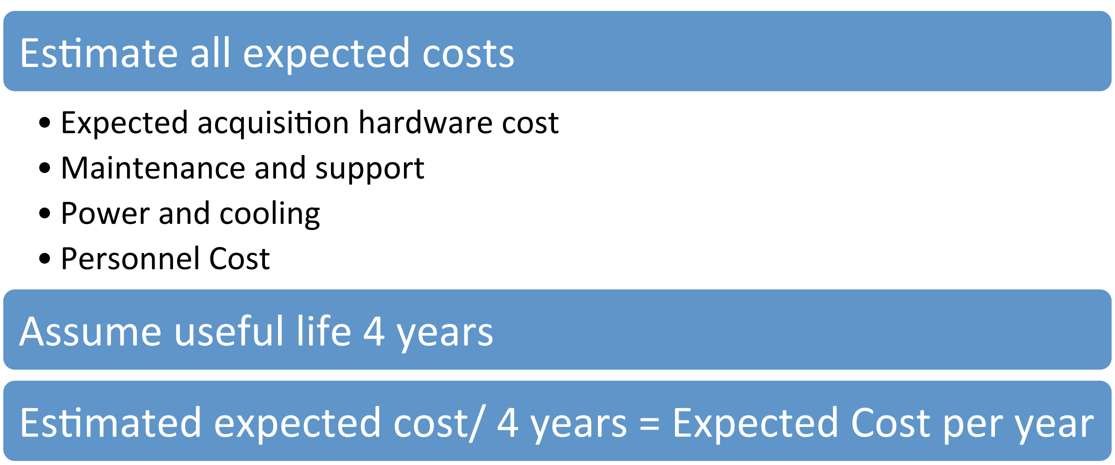


Figure 1: Graphical representation of a TCO case study

---

how to make cost estimates at each stage with reference to construction projects. The method includes costs consideration from the initial appraisal to the completion and post-occupation phases, including the disposal of the asset. http://ec.europa.eu/enterprise/sectors/construction/studies/life-cycle-costing_en.htm. Therefore Life Cycle costing uses the main principles of TCO.

[15] http://en.wikipedia.org/wiki/Total_cost_of_ownership

[16] http://en.wikipedia.org/wiki/Full_Cost_Accounting

[17] For an analytical review of the characteristics of the costing mythologies available please consult (Hilton, Maher and Selto, 2008).

[18] http://en.wikipedia.org/wiki/LRIC

[19] http://link.springer.com/content/pdf/10.1007%2Fs00799-012-0092-1

The **TCO model** has been used in literature (Nazir and Sorensen (2010), Walker (2009)) as a basis for calculating the costs of e-Infrastructures. TCO paradigm is a useful tool when the cost of a specific project is to be assessed over its useful life. The developers of such models adopt a *forward looking* stance covering all *expected* costs over the project's lifetime. This cost is divided by the anticipated useful life to come up with a yearly cost. However, in order for this analysis to be precise enough, several details have to be taken into consideration as basis of the future cost estimates. Such details are very difficult to gather consistently across several organisations. PRACE intends to follow a TCO costing methodology in order to assess the cost of contributing infrastructure[20]. A graphical representation of TCO is found in Figure 1.

On the other hand, **Full Cost Accounting methodology** is a typical costing methodology pertaining to all industries and sectors. It relies on actual cost accounting data information available within the cost accounting systems of organisations, i.e. it adopts a *backward looking* stance. Detailed *actual* data (in line item format) is attributed and allocated on the basis of specific algorithms and various costing procedures to services/products in order to come up with the "cost per unit" of the object being analysed. However, the reliability of the cost data depends heavily on the robustness of the cost accounting system. But even in the case of highly developed accounting systems, costs that are not funded by the organisation are not registered as parts of the total costs. FCA relies on actual cost accounting information (historical) available within the cost accounting systems of companies. Moreover, costs registered in accounting books are influenced by tax considerations (e.g. obsolete infrastructure not used any more would keep on being depreciated). Nevertheless, FCA is a very useful tool for organizations to control costs and assess performance. Information streaming from FCA can also assist pricing decisions. A graphical representation of FCA is found in Figure 2.

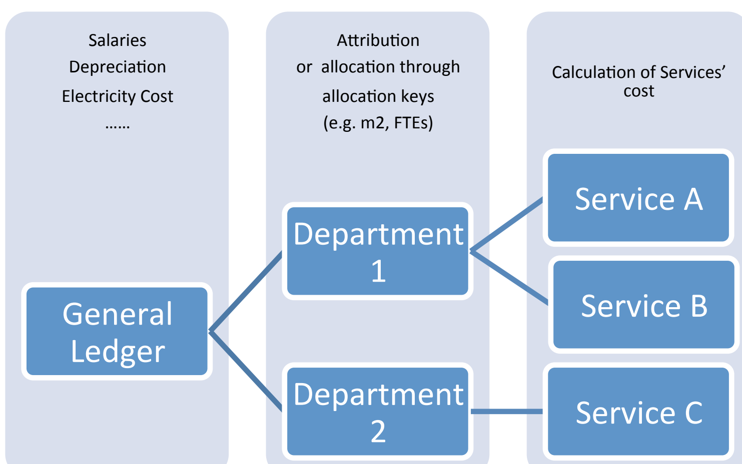


Figure 2: Graphical representation of FCA

[20] This exercise had not been completed by the end of January 2013. The cost categories intended to be included are presented in **section 3.**

2.2 Cost per core hour studies

The cost of execution of a process for an hour in a computer using a "core" in a processor has become a de facto reference. This is due to the limit reached in the frequency of the processors, and the increasing number of cores per processor. The advantage of using a single number for comparison between different resources is obvious, however the reader must be aware of the potential large diversity of the systems, with clear implications in the final performance of the user applications. Even if most of the HTC systems, and many of the HPC ones, are clusters with individual nodes using a dual processor server, the differences in cost and performance per core can be large. For example a "core" of a last generation processor running at 3 GHz with 8 GB/RAM memory per core, can have a cost up to ten times higher than a "core" of a processor running at 2 GHz with 2GB/RAM memory per core. This fact is reflected in the benchmarks' results.

Except for the studies that discuss methodological issues, there is one other category of papers and studies dealing with the cost per core hour under different settings. The ultimate goal of these studies is to **compare in-house costs with commercial cloud offerings**. A comprehensive example on this subject is presented in the Magellan final report (2011). Magellan is a project funded through the U.S. Department of Energy (DOE) Office of Advanced Scientific Computing Research (ASCR). Its goal is to investigate the potential role of cloud computing in addressing the computing needs for the DOE Office of Science (SC). In the conclusions of the project's final report it is stated, among others, that "the cost analysis shows that DOE centres are cost competitive, typically 3-7x less expensive when compared to commercial cloud providers". Hawtin et al (2012) in a cost analysis of cloud computing for research (on behalf of EPSRC and JISC) report that the more powerful cloud computing instances, rented on an hourly basis, appear to be one-and-a-half to two times more expensive per core-hour than well-managed, locally-provided clusters in modern data centres operating at high utilisation levels. However, other purchasing models (such as 'Reserved Instances') can reduce the costs to parity.

Smith in his thesis (Smith, 2011) presents a model for calculating the base cost for a core hour of computation in Purdue University's campus grid. With the cost model developed, the author analyses the benefits gained from using the grid, based on the number of hours of delivered, number of computations completed, and the number of users and faculty members served. In the same vein, a practical example of cost calculations and comparisons with cloud providers refers to Hyak shared compute facility at the University of Washington[21]. Carlyle et al (2010), present a case study of costs incurred by faculty end-users of Purdue University's HPC "community cluster" program. The authors develop and present a per node-hour cloud computing equivalent cost that is based upon actual usage patterns of the community cluster participants and is suitable for direct comparison to hourly costs charged by one commercial cloud computing provider. They find that the majority of community cluster participants incur substantially lower out-of-pocket costs in this community cluster program than in purchasing cloud computing HPC products. Another case study is found in Walker (2009). Walker proposes a modelling tool that can quantitatively compare the cost of leasing CPU time from online services to that of purchasing and using

[21] http://escience.washington.edu/get-help-now/hyak-operating-costs-and-comparison-commercial-alternatives

a server cluster of equivalent capability. Cohen and Karagiannis (2011) made a first attempt to approximate the cost of EGI. Their study, conducted within the e-IRGSP2 project, attempted an approximation of the cost of the EGI pan–European grid infrastructure for 2009 by extrapolating the cost of a few (seven) selected NGIs as example cases. Moreover, this study presents the cost per CPU core hour under alternative CPU utilisation rates and performs as a final step, rough comparisons between these costs and Amazon EC2 prices.

Risch and Altmann (2008) analyse the question whether using the Cloud is financially advantageous compared to the Amazon.com EC2 service. To perform this analysis they calculate the costs of computing resources in different usage scenarios, reflecting cases where Cloud resources and in-house resources are used. The comparison of the costs reveals that while the Cloud is cheaper in the short term, it is not a good investment in the long term and, thus, the existence of a Cloud economy will not lead to an end of ownership but rather to a reduction of in-house resources and more efficient resource usage.

The review of relevant literature briefly discussed above enabled us to take a more informed decision on two issues: a) the identification of the distinctive cost objects for which cost data should be gathered in order to run our costing exercise and b) the costing methodology that would best fit to our requirements (we discuss this issue in more detail in the methodology section). Moreover, we had identified the cost per core hour in several settings to compare with the results of our cost analysis.

2.3 In-house computing infrastructures and commercial cloud solutions

As part of the state of the art analysis, a number of comparison studies were evaluated to better understand the pros and cons of the in-house computing infrastructures vs. commercial cloud offerings. In terms of scope, the in-house computing infrastructures may refer to the European initiatives e.g. EGI, PRACE and other national or regional computing facilities. On the other hand, there are many commercial cloud providers e.g. Amazon EC2, Rackspace, Salesforce.com, etc. In the following few paragraphs, we summarise the analysis of various comparison studies that were included in the state of the art review.

One of the most common arguments made for the adoption of cloud computing is the potential cost saving compared to deploying and operating in-house infrastructure. This argument is based on several parameters. Commercial clouds consolidate demand across a large customer base, resulting economies of scale that small departmental clusters cannot achieve such as lower number of FTEs per core, stronger purchasing power when negotiating with the vendor, and better power efficiency since large systems can justify investing more in the design of the cooling infrastructure. Low upfront costs and pay-as-you go models are also considered advantages of clouds. Clouds allow users to avoid both time-investments and costs associated with building out facilities, procuring hardware and services, and deploying systems. Users are able to get access to on-demand resources and only pay for the services they use. Ultimately, whether a cloud offering is less costly than owning and operating in-house resources is very dependent on the details of the workload, including characteristics such as scaling requirements, overall utilization, and time criticality of the workload (Magellan, 2011).

Lin and Chen (2012) investigated how cloud computing is understood by IT professionals and the concerns that IT professionals have in regard to the adoption of cloud services. The findings of this study conducted through interviews in Taiwan suggest that while the benefits of cloud computing such as its computational power and ability to help companies save costs are often mentioned in the literature, the primary concerns that IT managers and software engineers have are compatibility of the cloud with companies' policy, Information Systems (IS) development environment, and business needs. The findings also suggest that most IT companies would not adopt cloud computing until the uncertainties associated with cloud computing, e.g. security and standardisation are reduced and successful business models emerge.

Marston et al. (2011) discuss cloud computing from several fronts. They identify the strengths (e.g. reduced infrastructure costs and energy savings as well as reduced upgrades and maintenance costs), weaknesses (e.g. the loss of physical control of the data that is put on the cloud), opportunities (e.g. energy consumption and carbon footprint reduction) and threats (e.g. lack of standards that may force customers into locked, proprietary systems that will gradually cost more and more over time) for the cloud computing industry. They also discuss information systems (IS) policy issues which cover a broad set of topics, from data privacy and data security to data ownership and audit.

Hammond et al. (2010) discuss cloud computing for research in the areas of compute and storage in relation to Infrastructure as a Service (IaaS) and Platform as a Service but not Software as a Service. Authors argue that critical thinking is still required from researchers and institutions as to what data storage or compute solution is most appropriate given functional requirements, budget, security, reliability, trust, etc. as well as the cloud services currently on offer.

The EGI (2011a) report aims to evaluate technologies such "Infrastructure as a Service', 'Platform as a Service' and 'Software as a Service', understand how they relate to EGI, and build a foundation for the integration of cloud and virtualisation[22] into the European production infrastructure. The EGI (2011b) report includes a cost analysis and comparisons to current market offers. The EGEE (2008) report lies within the same realm. This report compared grid and cloud computing services, taking a practical look at implementations of the Enabling Grids for E-sciencE (EGEE) project for grid and the Amazon Web Service (AWS) for cloud. Taking performance, scale, ease of use, costs, functionality and other aspects into consideration, the report looked at the overall opportunity that converging cloud and grid services can bring to users.

Misra and Mondel (2011) on the other hand propose a model to assist companies by analysing several characteristics of their own business as well as pre-existing IT resources to identify their favourability in the migration to the Cloud Architecture. They also develop a general Return on Investment (ROI) model that embodies various intangible impacts of Cloud Computing, apart from the cost, in order to provide a tool for a broader perspective and insight cloud computing decision assessment.

[22] EGI (2012) discusses in more depth EGI.eu strategy towards virtualization (http://go.egi.eu/EGI2020). The document describes the EGI strategy to evolve into a universal federated platform for supporting compute and data intensive Research and Education communities.

The pricing of cloud services is also another issue. The e-Science Institute[23] focuses on the steady decrease in prices offered by Amazon for their services. They comment that over time, the price to rent one unit of resources for three years of continuous usage has fallen dramatically as Amazon offered new instance types, offered new long-term pricing plans, and lowered prices outright across the board.

2.4 Business models

IT is an ever-evolving landscape; today countless innovative business models are emerging as well as other entirely new industries forming business models around cloud computing, social media and smart devices. This growing trend is now spilling over into the public sector as well. Funding agencies across Europe are looking to focus investments on more targeted areas that will stimulate innovation and economic growth.

Defining business models help provide the necessary structure for organizations to develop a series of organisational and strategic components for delivering its value to its consumers. However, a common misconception in the public sector is that a business model is how to generate revenue or conduct commercial activities. While any organisation that creates and delivers value must be able to generate enough revenue to cover its expenses, a business model is much more than that. It can be basic or complex, but is simply "the rationale or description of how an organisation creates, delivers, and captures value". Osterwalder (2010)[24] developed a simple one-page template to discuss the key elements on business models, and cost and revenue streams are two important elements.

In the following sub-sections the state of the art in business models for publicly-funded federated e-Infrastructures is reviewed, along with some basic trends in the business and pricing modes of cloud computing.

2.5 Federated Infrastructures and Business Models

Publicly funded e-Infrastructures are inherently different in nature both in terms of funding structure and purpose. The main focus of e-Infrastructures is to support and facilitate world-class research. The e-Infrastructures themselves have been built, maintained and operated through public financing, which comes with certain restrictions on what can and cannot be done with the resources (e.g. selling computing capacity, running commercial applications).

Therefore, the application of business models drastically changes in the public sector from that in the commercial world. e-Infrastructures in Europe operate within a publicly funded research and academic environment providing services free at point of delivery with resources bought from grants dedicated to certain groups or disciplines either by direct allocation or by peer review. With the advent of cloud computing, business models and user expectations are shifting towards on-demand provision increasing flexibility and agility. In a distributed environment, services are provided by a variety of different organizations spread across Europe and beyond.

[23] http://escience.washington.edu/blog/cloud-economics-visualizing-aws-prices-over-time

[24] A. Osterwalder "Business Model Generation", Wiley 2010

Over the last few years, EGI has been investigating this as part of their sustainability strategy and is being picked up in other project efforts when discussing business models and service provision and management for e-Infrastructures (FedSM[25]). In a federated environment such as e-Infrastructures, the concept of brokerage applies well. Thus the definition of specific roles and functions is needed and generic brokerage models need to be evaluated. According to FedSM, the following roles are identified:

- the Federator, which provides the technology, processes and governance to enable access to an integrated set of services from autonomous organizations,
- the Resource Provider, which offers access to ICT resources through service abstractions (e.g., computing power, storage),
- the Customer who negotiates the level of services and commissions the service provider or broker and may pay, doing so on behalf of a number of consumers (users)
- and the Consumer, the person actually using the service (user).

There are a number of different brokerage models used by a variety of domains. In this section, we present three main models that can be applied in e-Infrastructures.

Model 1: Independent Advisor

In the 'Independent Advisor' model, the federator provides a general listing of services, facilitates relationships between customers, consumers and resource providers while playing only a support role if required during the service lifecycle. Through the federator, resource providers can promote their services to customers, while retaining the direct dialogue concerning the resource allocation, contracts and financial transactions. This model requires the customer to interact within individual (potentially multiple) resource providers to obtain the services it requires (Figure 3). Therefore, interactions are decentralized leading to higher overheads for the number of relationships that customers/resource providers must maintain. The federator is able to fund the services it provides through a membership model, which restricts the customers and resource providers that can use them. This is the current model used by EGI.eu.

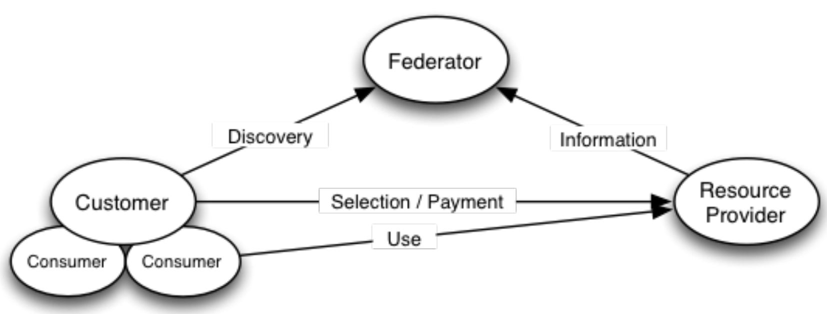


Figure 3: Independent Advisor Broker Model

[25] FedSM D3.1 - Business Models for Federated Infrastructures. http://www.fedsm.eu/sites/default/files/FedSM-D3.1-Business_models-v1.0.pdf

Model 2: Matchmaker

In the 'Matchmaker' model, the resource allocation is managed by the federator. The customer discusses requirements and receives a resource allocation from the federator with a resource provider. The contractual agreement is established by the federator with the customer on behalf of the resource provider but any financial transaction is handled directly between the customer and resource provider (Figure 4). The resource provider pays the federator for establishing the contractual agreement. This model is more suitable for customers who need access to many resource providers.

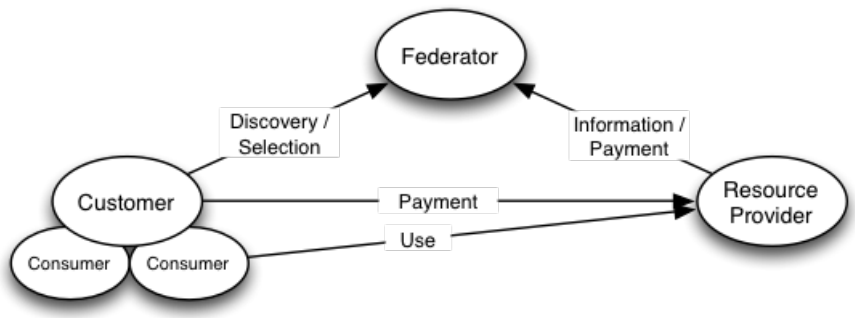


Figure 4: Matchmaker Broker Model

Model 3: One Stop Shop

The 'One Stop Shop' model fully relies on the federator to handle the service publication, matchmaking, contract and agreement negotiation, as well as financial transactions (Figure 5). The resource provider receives payment for the resources used by the consumer through the federator. Reliance on such a service reduces organisation overhead on both customers and resource providers by offering them a single point-of-contact to many independent counter-parts.

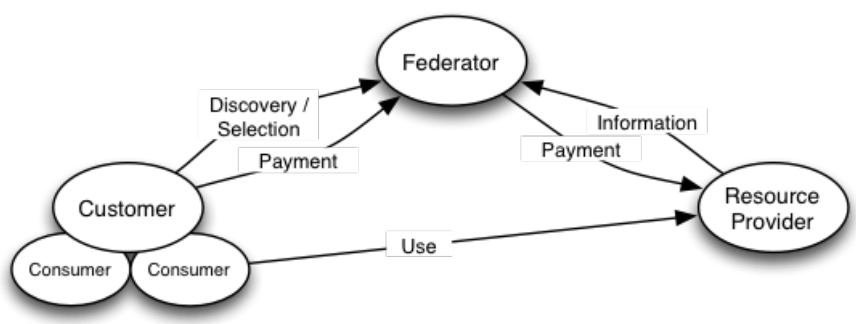


Figure 5: One Stop Shop Broker Model

2.6 Commercial Clouds and Pricing Models

In the IT world, cloud computing not only offers the opportunity for a wide range of products and services but a number of different business models that can be offered. Based on the product or services various pricing models are applied. With cloud computing, three standard service categories emerged: Infrastructure as a Service (IaaS), Platform as a Service (PaaS) and Software as a Service (SaaS). A few examples of some business models in cloud computing has been the industry standard Amazon Web Services offering IaaS, basically selling its extra computing capacity as a service (e.g. computing power, storage). Over the years, this has grown in complexity in terms of both services and pricing models. Dropbox provides an online storage facility for more daily users, which then applies a Freemium based model (i.e. 2 GB storage free, +X is paid).

Osterwalder (2010) classified the pricing models around fixed (predefined based on static variables) and dynamic (based on market conditions). Harmon et al[26] (2009) proposed a factor-based pricing mechanism (cost + value). Paleologo (2004)[27] observed that in on–demand service environments such as cloud computing traditional pricing mechanisms such as cost plus pricing may be inadequate. This is due to the dynamic factors of cloud computing such as shorter contracts, reduced switching costs, customer lock-in, uncertain demand, and shorter life cycles etc. Denne[28] (2007) elaborated on advanced ways to implement pay per use pricing mechanism, which is also currently being investigated by EGI.

2.7 Green IT – Energy Costs

Green computing, as it is now often referred to, seeks to minimise the environmental impact of computing through different methods, for example:

- Improving the efficiency of computing centres
- "Follow-the-moon" approach that migrates energy-intensive calculations to night time of the centres
- Choosing areas where low CO2 electricity is available for computing centres
- Using specialised hardware (e.g. power sources, transformers, lower CPU clock speed)
- Using the heat generated by the computers to heat offices or residential buildings
- Recycling the hardware
- Using IT to reduce emission in other operations (e.g. reducing travel through telecommuting)

As the list above shows, "Green IT" is a very wide topic, incorporating extremes like minimisation of the so-called e-waste to developing scheduling algorithms that run the tasks in an optimal way (Niemi et al. 2012).[29] Generally, energy and resource optimization in scientific computing has mostly focused on hardware and infrastructure issues, for example, the development of more efficient hardware or optimizing cooling of computer centres. Recently, they had been focusing also on operational methods such as workload management and operating systems and application software optimization for energy-efficiency. There are also a few studies focusing on how increased memory utilisation affects throughput and in particular energy consumption in scientific computing (Niemi and Hameri, 2012). The most well-known method for comparing energy efficiency of data centres is Power Usage Effectiveness (PUE)[30] metric. This is a ratio of the total facility power / IT equipment power. It indicates how much of the energy is lost in cooling, power distribution, and other infrastructures. The benefit of this measure is that it is feasible to collect this information in a large-scale survey in a consistent manner that allows comparison between different centres or regions. However, in addition to not indicating how the efficiency of operations is achieved, some Green IT approaches do not have a direct impact on PUE.

---

[26] Harmon, R., Demirkan, H., Hefley, B. & Auseklies, N. (2009). Pricing Strategies for Information Technology Services: A Value-Based Approach. Proceedings of 42nd Hawaii International Conference on System Sciences, pp. 1-10.

[27] Paleologo, G. (2004). Price-at-Risk: A methodology for pricing utility computing services. IBM Systems Journal, Vol. 43, No. 1, pp. 20-31.

[28] Denne, M. (2007). Pricing utility computing services. International Journal of Web Services Research, Vol. 4, No. 2, pp. 114-127.

[29] http://www.igi-global.com/chapter/improving-energy-efficiency-scientific-computing/67303

[30] Power Usage efficiency (PUE) is determined by dividing the amount of power entering a datacentre by the power to use to run the computer infrastructure within it.

Within the broad area of Green computing research, we contained our review in papers dealing with **electricity cost** and cost of data centres in general. The electricity efficiency issues fall within the broader area of Green IT. In Koomey et al. (2009) the relationship between the processing power of computers and the electricity required to deliver that performance is analysed. Rasmussen (2011) proposes a method for measuring total cost of ownership (TCO) of data centre network room physical infrastructure and relates these costs to the overall information technology infrastructure with examples. In several cases, PUE has become a measure of datacentre efficiency. Cordis (2011) discusses potential actions that could contribute towards improving efficiency in data centres and reducing costs. The report presents a discussion about ways that Europe would be mobilised to minimise both the environmental and economic impact of such centres, while addressing the issue that the amount of data shows a considerable increasing trend. The strategies are discussed under two perspectives: The "low hanging fruit actions" and the long term programs. The report touches the whole resource efficiency consideration of data centres (energy consumption, thermal dissipation, use of water for cooling, carbon footprint and the construction of data centres per se). The report considers critical the measuring of impact through metrics welcoming PUE as a first step, while acknowledging differences among data centres. Actions that can contribute towards decreasing power consumption include among others, applications software efficiency, hardware/software co-design, advances in processor design, new approaches in minimising cooling requirements, improvements to uninterruptible power supplies (UPSs) and usage of renewable energy sources. Following Cordis (2011) we consider PUE as a significant metric to assess awareness regarding Green IT.

According to a recent study in the UK comprising 27 UK colleges and universities (Hopkinson and James, 2012), the total energy for all PCs (including desktops, laptops, thin clients and monitors) ranged from 144 kWh/year to 587 kWh/year. Especially in HPC-intensive universities the average energy consumption per PC is 309 kWh/year.

Finally, it is worth mentioning that several computing centres that participated in the e-FISCAL study reported (through the survey) that they have embarked into interesting Green IT initiatives.[31]

[31] See for example http://www.rehva.eu/en/608.space-heating-with-waste-heat-from-computer-centre-in-the-vattenfall-head-office, http://www.csc.fi/english/csc/news/news/data and http://www.pdc.kth.se/news/news-repository/excess-heat-from-pdcs-supercomputer-keeps-kth-building-warm

3. Methodology

The e-FISCAL project has applied the methodology discussed in this section in order to achieve the project's goals. As the main focus of the project lies on costing issues the graph that follows concentrates on the development of the costing model and its operationalisation in order to come up with sufficient data to answer the first three research questions (i.e. Research question 1: To analyse the cost structure of HPC and HTC centres, Research question 2: To approximate the cost of the HPC and HTC in Europe and Research question 3: To compare the cost of HPC and HTC centres with commercial Cloud Offerings). However, along this procedure input useful to address the secondary two research questions (i.e. Research question 4: Identification of suitable business models and research question 5: An analysis of the Green IT initiatives in HPC and HTC centres) is gathered.

The methodology comprises six stages that are presented in the graph below and are discussed in the rest of the methodology section:

- State of the art review
- Development of the generic cost model
- Sample selection
- Questionnaire development
- Questionnaire dissemination and follow up
- Analysis of data and conclusions

A graphical representation of the sequence of these steps is shown in Figure 6.

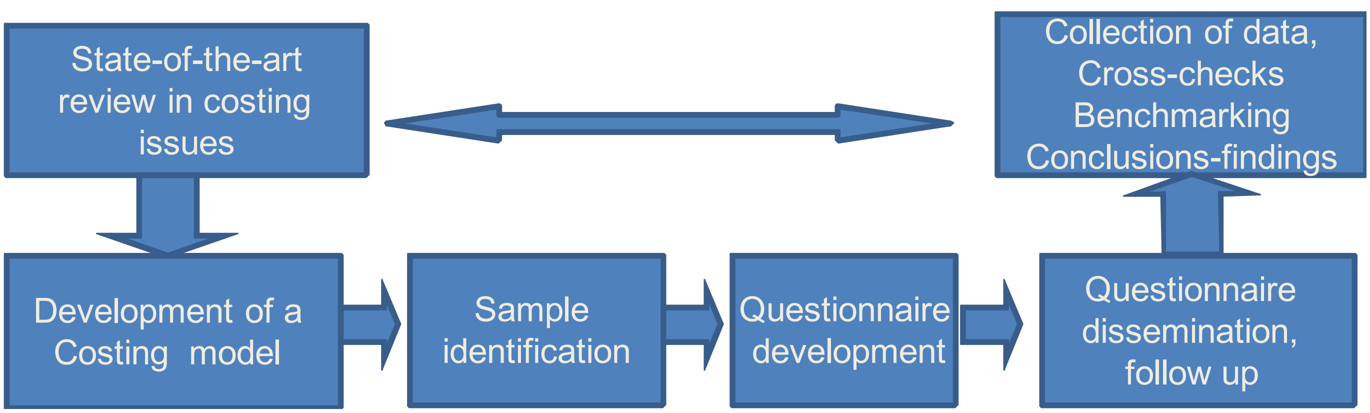


Figure 6: Overview of the proposed methodology

3.1 State of the art review

The state of the art has been analytically discussed on Section 2. In summary, this methodological step comprised a detailed review of the relevant literature. The review covered academic papers, industry project results, EU studies' results, vendor analyses and relevant studies to acquire an overview of the current state of the art relevant to the scope the project. This first step had multiple orientations including the identification of relevant research works that dealt with e-Infrastructure costing issues in general (Afran and Bancalore, 2007; Walker, 2009; Nazir and Sorensen, 2010; Opitz et al, 2008), electricity and premises costing (Koomey, 2008; Jie et al., 2011), business models applied by e-Infrastructures (EGI-InSPIRE EU deliverable D2.7, 2011), papers that compare cloud and grid options (Foster et al., 2008; Kondo et al., 2009) or discuss cloud computing issues (Klems et al.,

2009; Hammond et al., 2010; Microsoft, 2010; Kim et al., 2009), papers that discuss the migration to the Cloud (Misra and Mondal, 2011), industry benchmarks (Crosswell, 2010), case studies (Hyak operating costs[32]) etc.
All papers reviewed are found on the project's state-of-the art repository (http://www.efiscal.eu/state-of-the-art). The repository contained as of 31/1/2013 50 references.

3.2 e-FISCAL costing model characteristics

The thorough state of the art review provided useful input to develop the costing model in at least two dimensions. Firstly, we had to decide on the cost categories to be included in the analysis and secondly on the costing methodology. As for the cost categories we decided to abide by the cost categorizations proposed by Optiz et al. (2008). This cost categories could be considered as standard in relevant cost studies. As for the costing methodology, we built a hybrid costing model that capitalises on Total Cost of Ownership (TCO) and the Full Cost Accounting (FCA) methodology properly adjusted to the characteristics of the project.

3.2.1 Costing methodology

The hybrid methodology developed for the scope of the project is suitable to calculate the total yearly cost of computing e-Infrastructures. Its materialization requires a two-step process: a) Simulation of the physical infrastructure and b) Development of the financial model.

a) Simulation of the physical infrastructure: This corresponds to the investment cost of the infrastructure which is approximated by taking into account the capacity in terms of cores, of storage devices, of interconnection devices and of auxiliary equipment and the actual purchase values of these elements.

b) Development of the financial model: The financial model is based on two pillars. The annualised cost of the simulated physical infrastructure and the operating cost of the physical infrastructure.

- The annualized cost of the simulated physical infrastructure is derived by applying depreciation rates to annualise the cost of the physical infrastructure simulated in the first step. This amount accounts for the yearly CAPEX (Capital expenditures).
- The operating cost of the physical infrastructure corresponds to the yearly costs for running the site/centre. In order to approximate the running cost, information about the operating costs is gathered per cost category (e.g. personnel costs, electricity, premises cost, etc.). This amount accounts for the OPEX (operating expenses).

More specifically, the **cost categories** that are included are as follows:

1. *Computing and storage hardware costs*. This category refers to CPU cores (and corresponding boxing), storage devices (disk and tape storage facilities) and related interconnect equipment (network devices). Service support costs fall in this category.
2. *Auxiliary equipment costs*, that correspond to the investment in cooling devices (air or liquid cooling), UPSs, power generators, power transformers, etc.

[32] http://escience.washington.edu/get-help-now/hyak-operating-costs-and-comparison-commercial-alternatives

The costs of (1) and (2) are used to simulate the physical infrastructure and they are then annualised through depreciation rates to come up with yearly CAPEX.

3. *Software costs* correspond to yearly costs for licensing and purchasing operating systems, middleware, support contracts, applications, 3rd party software, compilers, etc.
4. *Personnel costs* relate to the salaries and all extra burdens assigned on salaries of all related categories, such as administrators and operators, middleware and application developers, trainers, dissemination persons, policy makers, managers, etc.
5. *Site operating costs* correspond to rentals or depreciation costs of hosting premises.
6. *Electricity cost* refers to the power usage for the consumption of the in-house network devices like servers, routers, etc. It also includes the electricity consumed for cooling and hosting.
7. *The network connectivity costs* correspond to the yearly cost for having connection to the internet.
8. *Other costs.* In this category all costs not falling in any of the previous ones should be registered. Personnel training costs, training certifications, travelling expenses – participation to conferences costs, fees paid to the university/institute for hosting the site/centre, insurance fees, even fees for cloud services could be registered.

The costs of (3) to (8) refer to the operating cost of physical infrastructure. They constitute the yearly OPEX.

The collection of information referring to both CAPEX and OPEX could permit, among others, the execution of a cost breakdown analysis and economies of scale assessments. Additionally, external data (e.g., from EGI, PRACE, market or other literature sources) can be gathered and used for verification and cross-checking.

The yearly cost of e-Infrastructures according to the scope of the study should be calculated for years 2010 and 2011. A graphical overview of the e-FISCAL methodology is presented in Figure 7. In the figure, we also indicate the origin of the data to be included in the model. It should be commented that the cost categories used in e-FISCAL coincide with the cost categories to be used by PRACE in its TCO analysis[33].

[33] TCO analysis for PRACE that into account the following:
Investment costs taking into account the expected lifespan of supercomputers, including installation costs; related Information Technology ("IT") equipment required for the operation (storage system, back-up and internal computer centre networks); buildings; technical facilities, including cooling, power supply; Maintenance of the supercomputers and related IT equipment and software licenses, including vendor support for hardware and software; Maintenance of the buildings and technical facilities; Electricity charge, including the depreciation cost of the power line and main substation if needed; The staff, including management, computer centre operation, building and technical infrastructure support; Changes and upgrades that might be required during the first five years (http://www.efiscal.eu/files/presentations/samos/PRACE-presentation%20ALG%20e-fiscal%2020120703%20V2.pdf).

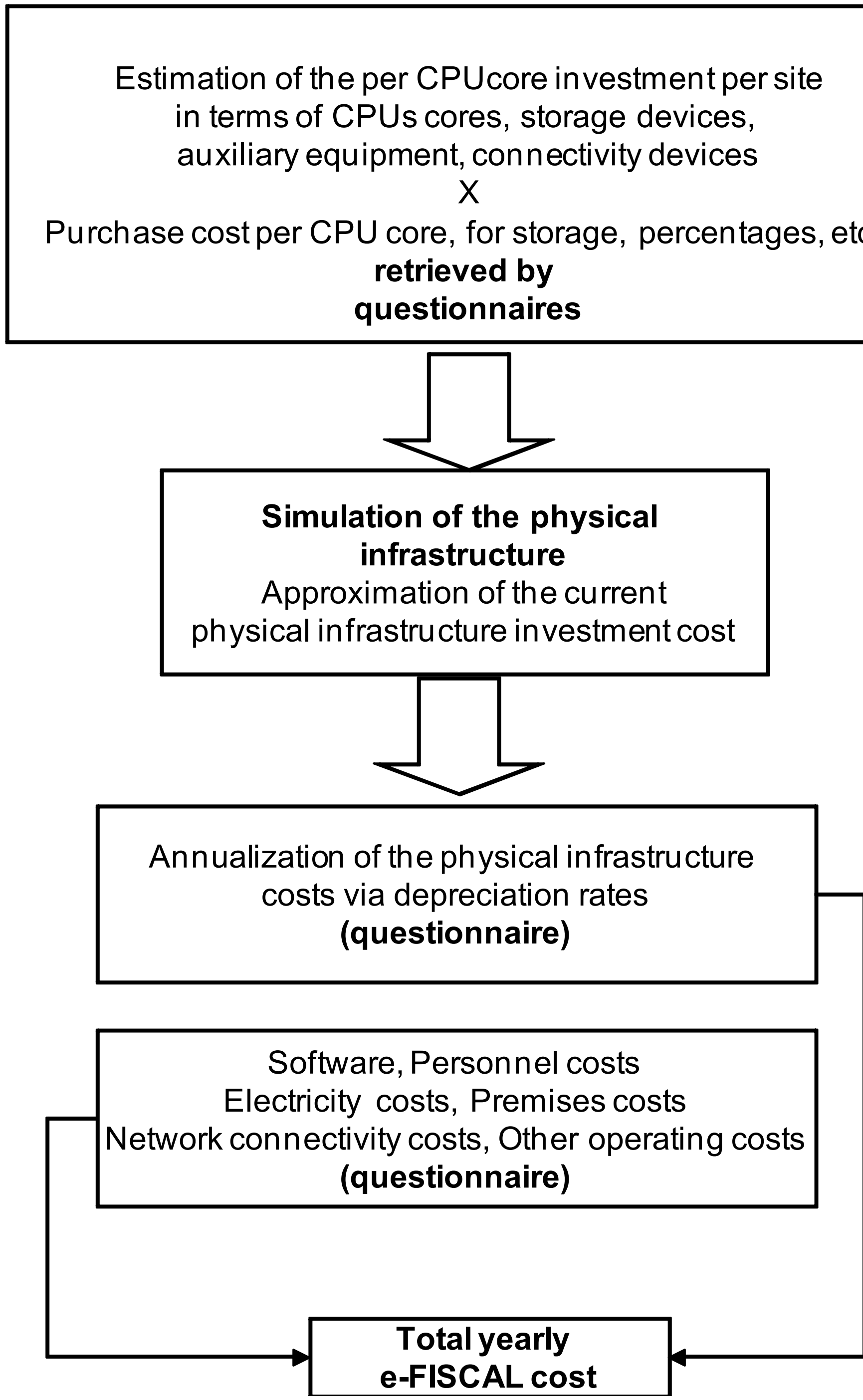


Figure 7: Overview of e-FISCAL cost methodology to calculate total yearly cost

As already mentioned above, the e-FISCAL methodology is a hybrid that capitalises on the characteristics of full cost accounting and TCO. Most importantly, it simplifies data collection that is a prerequisite in order for the e-FISCAL project to be able to gather the necessary information in order to make the costing approximations. As the researchers of the e-FISCAL project are external to the sites/centre entities with no access to accounting books and detailed information, the methodology developed balances accuracy and easiness of data reporting. A graphical presentation of the proposed methodology compared to full cost accounting and TCO is given in Figure 8.

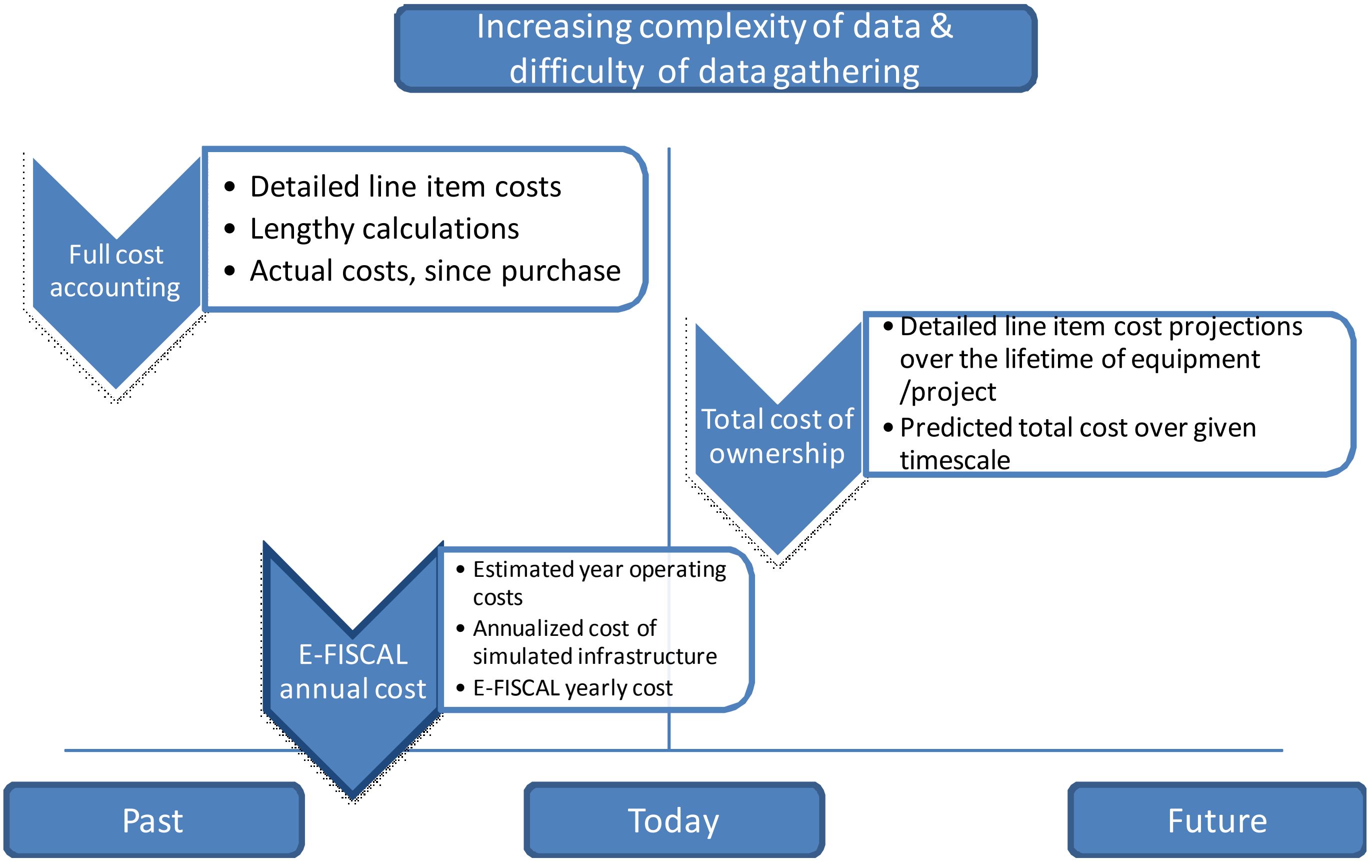


Figure 8: Comparing TCO, FCA and e-FISCAL methodology in cost calculations

As it can be inferred by the graph above Full Cost Accounting may end up using obsolete assumptions, whereas TCO requires making assumptions on the whole lifecycle of the equipment (which in the case of machine rooms can be decades), bringing in their own uncertainties. Our proposed hybrid methodology is suitable for estimating the level and breakdown of current costs as well as allowing for the accurate projection into the short-term. The approach balances the level of detail and the accuracy of necessary costs against the effort required to gather and report the cost data. Moreover, it accounts for costs incurred but not paid by the site. As Intersect 360 comments in both reports of 2012 and 2013, several of their respondents stated that either the facilities cost were not in their budget or that they did not pay for every item (such as power or building/floor space) in their budget.

3.3 Research setting

In order to be able to approximate the overall cost of the dedicated European HTC and HPC infrastructures we would ideally need cost information from the total population of dedicated European HTC and HPC infrastructures. However, as it is in practical terms very difficult to get data from every single HTC or HPC centre (i.e. to run a census), our analysis is expected to come up with robust and reliable results if there is an acceptable rate of participating entities. Having said that, it is not the absolute percentage of contributors that matter the most; their **diversity** is what it counts more. In order to be able to extrapolate the findings to the whole population of dedicated European HTC and HPC infrastructures, an adequate representation of sites/centres in different countries, of different size and of different nature (HTC/HPC) is needed. Getting input from sites/centres of different size would also permit the identification of any economies of scale effects.

In order to cover as many cases (sites/centres) as possible, the method to collect data has been decided, as early as in the proposal phase, to be that of the questionnaire. The use of questionnaires permits the collection of information in a homogeneous and standard way and is less demanding in resources compared to interviews. Also, as the cost data asked is highly likely not in the possession of only one person but needs the cooperation of several people from different posts, the use of questionnaire facilitates fragmented provision of information as information is registered in a given document at any time convenient to the respondent.

3.4 Survey Instrument

The developed questionnaire for the scope of the study was designed to provide the relevant inputs to satisfy requirements of the cost model presented in section 3.2.1.

The questionnaire design needed to balance between conciseness and the ability to accurately distinguish important differences in cost structures of the centres. To strike an optimal balance, different versions of the questionnaire were tested internally in several iterations, after which the final version was sent to a small set of voluntary participants in the survey.

The survey instrument went through an intensive review process that could be summarized as follows:

- 1st Draft questionnaire (already in August 2011)
- Internal review and feedback from all e-FISCAL partners
- 2nd Draft questionnaire
- Presentation of the questionnaire to the dedicated e-FISCAL workshop in EGI Technical Forum in Lyon (21/9/2011)[34]
- Collection of both on-site and off-site comments from participants
- Pre-final questionnaire (October 2011)
- Pilot testing of the questionnaire (1 HTC site, 1 HPC centre and one mixed HTC/HPC centre
- Final questionnaire integrating the comments from the above centres (November 2011)

The review process performed concluded on a survey instrument that:

- Balances the effort in gathering and reporting information with the usefulness of the information retrieved
- Ensures that all necessary data for running the costing model is requested
- Ensures that information available by other sources (e.g. EGI portal) is used to avoid asking respondents the same questions and to avoid duplication of work
- Includes questions that are both clear and easily understood in a unilateral way from all respondents
- Is applicable to both HTC and HPC centres

[34] A copy of this version of the questionnaire is found in https://www.egi.eu/indico/conferenceTimeTable.py?confId=452#20110921.

The final version of the questionnaire that is found in Appendix 7.1 had two main sections. The first covers the necessary input data for the calculation of the total yearly cost for both years 2010 and 2011 (i.e. amortized investment costs and operating expenses). Therefore, there are questions referring to the investment in e-Infrastructure elements (e.g. hardware such as computing, storage, network and auxiliary equipment) as well as operating expenses related questions (e.g. personnel, electricity, premises costs). The second section is related to the sustainability outlook and Green IT aspects, where questions about the current and future use of services by commercial service providers are discussed. The questionnaire is common for both HTC and HPC centres.

Moreover, in the first section, the questionnaire follows a thematic structure (i.e. each cost category is analysed in detail in a sequential order). This structure is justified on the grounds that it permits the respondents to concentrate on all aspects that are related to each cost category at a time. Moreover, within each cost category there are questions related to the collection of both monetary values and metric volumes (e.g. average cost for CPUs acquisition and number of CPUs). The collection of data under this dual format facilitates both the execution of sensitivity analysis as well as the calculation of benchmarks (e.g. cost per CPU). In some cases, the structure of the questions allows the respondents to present the value of a cost element as a percentage of another cost element when detailed information is not available (e.g. auxiliary equipment cost as percentage of computing and hardware storage capacity acquisition cost).

The questionnaire is administered using a web interface (Survey monkey). The on-line version of the questionnaire can be found in http://www.surveymonkey.com/s/e-FISCALquest. A portable document format (PDF) version is found in Appendix 7.1. Also, an-excel based version of the questionnaire is available on Appendix 7.2 and on http://www.efiscal.eu/tools.

3.5 Survey instrument dissemination and follow up

The survey instrument dissemination and follow up has been planned from the beginning of the project in order to abide by methodological standard practices. The strategy was presented in the Lyon workshop on September 2011. The dissemination strategy was designed in such a way to achieve as much participation of the dedicated European HPC and HTC centres to the e-FISCAL project as possible through providing cost data. Also, the strategy aimed at providing continuous validation of the input data. The validation process was planned to be performed through a dual lens. Firstly, by constantly analysing input data to identify any inconsistencies, outliers or possible misunderstandings that would result in data deviating from what seems to be a common standard from the respondents' point of view. Secondly, by comparing input data with existing benchmarks found in market studies and vendor reports to determine any significant variances.

The process that was decided to be followed was to prepare written clarification questions to all respondents to be sent by e-mail, when necessary, and/or communicate with them either in teleconference sessions or in the workshops organised by the e-FISCAL project. Especially through the interviews conducted in the workshops, we would be able to identify useful and detailed information that would assist us in better understanding the cost structures and cost behaviour issues in e-Infrastructures. Eventually, we have sent written clarification questions to 27 respondents of which we took 19 answers back and we performed 8 in depth interviews during the Summer Workshop (July 2012) and the Final Workshop (January 2013).

An overview of the dissemination and follow up strategy in graphical form follows:

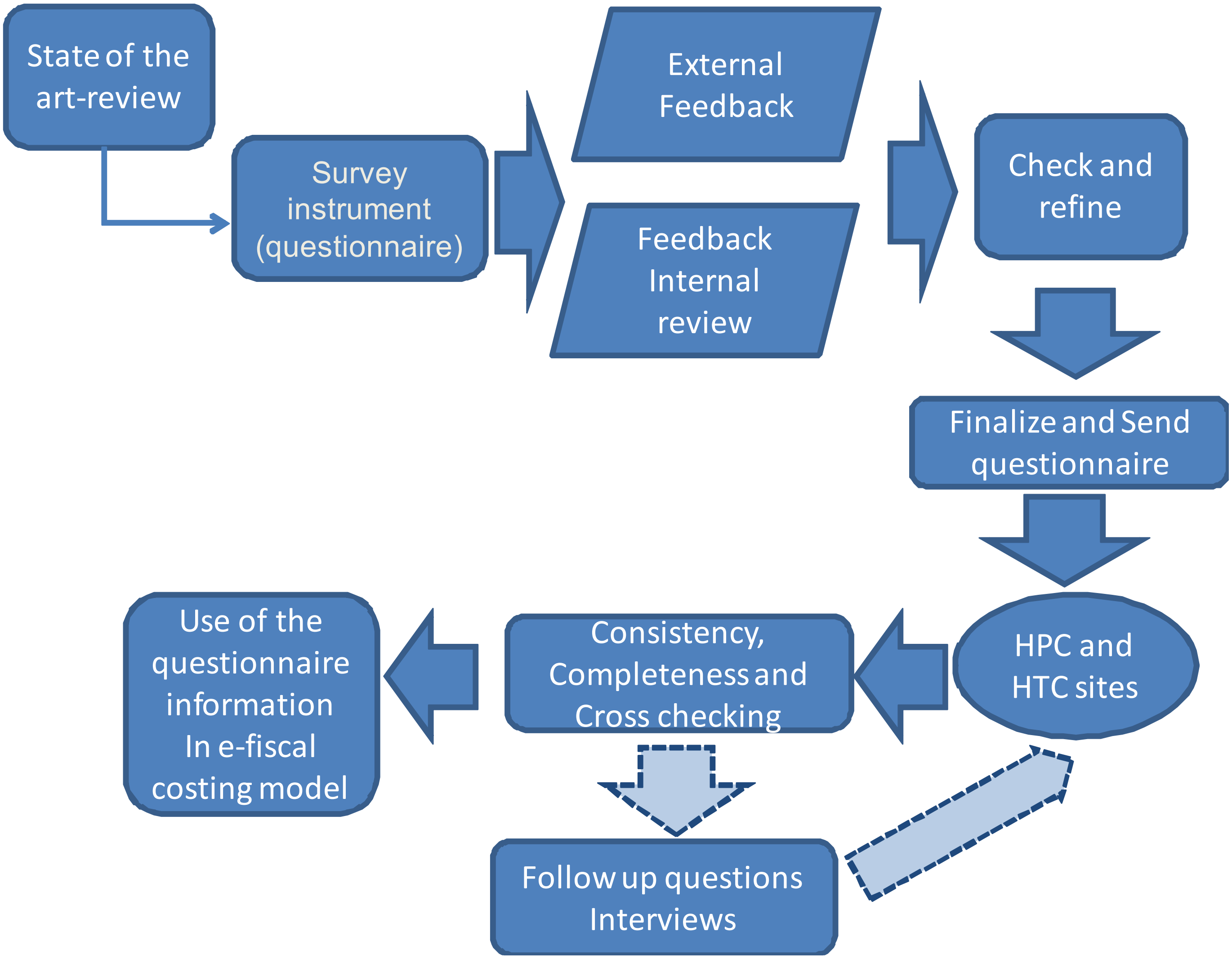


Figure 9: Dissemination and follow up strategy

The questionnaire was presented at the NGI International Liaison kick-off meeting in Amsterdam to inform the official NGI contacts about the existence of this initiative and the near-term announcement of the questionnaire (https://www.egi.eu/indico/event/659).

The project made a press release to announce the initiation of the survey[35]. The dissemination of the questionnaire was done on the week of the 5 December 2011. The press release was launched through several channels (such as Cordis Wire and AlphaGalileo), and was picked up by publications HPCWire, iSGTW, RoadRunner[36], DallasNews[37] and OSSREA[38], among others. At the same time, the questionnaire was promoted by e-mails to personal contacts.

[35] The announcement is also found on the e-fiscal site (http://www.efiscal.eu/node/15).
[36] http://www.rr.com/
[37] http://topics.dallasnews.com/

The survey instrument was available in both on-line format (SurveyMonkey software[39]) as well as a pdf editable format. As stated on the invitation, the questionnaire was soliciting answers from NGIs, national HPC coordinators, as well as from individual HTC/HPC centres. This was deemed necessary as some costs or related information is kept only at this level (e.g. energy consumption or housing costs, especially if a national entity is not paying for these).

The initial deadline for questionnaire completion was the end of January 2012. However, an extension was given for the end of February due to several requests. The extension was officially announced on the project web site (http://www.efiscal.eu/survey-extension). In parallel with the extension, a letter from the European Commission GÉANT and e-Infrastructures Unit Head was sent to the EGI and PRACE Councils[40,41] along with the e-IRG delegates[42], putting the relevant weight and also explaining that dedicated effort for filling in a demanding questionnaire would have been non-economic. The letter was catalysing besides specific upscale HPC centres (such as the PRACE Tier-0 centres) that had binding non-disclosure agreements with the hardware vendors and could not reveal their costs.

For questions related to the questionnaire, a dedicated e-mail address was set up (survey@efiscal.eu). Additionally, respondents could send their queries through the contact page of the project. Through informal discussions as well as through e-mail exchange, it was evident that several people had concerns about the confidentially of their responses as well as the potential use of the information provided for purposes outside the scope of the study. In order to dissolve any doubts about the confidentiality of input data, we developed a sample presentation with dummy figures in order to show how we intended for the results to be reported[43]. In this presentation we made clear that the data provided will be processed in such a way that it will not be possible to identify the provider of any type of information. We also developed a more formal document outlining the confidentiality practices of the project that could be used as a basis for a signed agreement between the project and the party providing the information. This document was never eventually used.

Finally, in order to stimulate participation and to ensure that the project workshop in Samos (aiming to present this report, meet with the respondents, interview them and receive their feedback) would gather maximum number of financial experts, we pledged that those being the first to fully complete the questionnaire would be invited to the e-FISCAL financial experts' workshop in summer 2012. Additionally, they would be entitled to have their travel expenses to attend the event covered.

---

[38] http://www.ossrea.net

[39] SurveyMonkey: www.surveymonkey.com

[40] EGI Council members: http://www.egi.eu/about/EGI.eu/council_members.html

[41] PRACE members: http://www.prace-ri.eu/Members

[42] e-Infrastructure Reflection Group members: http://www.e-irg.eu/about-e-irg/members.html

[43] This presentation if found in http://www.efiscal.eu/files/sample%20publication%20eFISCAL.pdf.

In an effort to get more inputs from more countries and centres during the second period of the project two new questionnaires were prepared. The first one was a user-friendly spreadsheet version based on the initial one and the second was a high-level one based on ranges (mores suitable for centres with confidentiality issues). The questionnaires were further disseminated and the project received four new questionnaires; two of which were new (from new sites and new countries) and two were updates from the previous ones.
Finally, the excel version of the questionnaire acted as basis for the on-line web-based tool.

3.6 Benchmarking exercise

As shown in the last step of our methodology, we provided for the execution of a benchmarking exercise. The objective of the benchmarking effort (also described in the D2.2) has been to carry out a performance comparison of different systems with identical specifications, while focusing on the HPC, HTC and Cloud infrastructures. The output of the benchmarking exercise would be used in the comparison between the prices of cloud providers and the costs derived from the e-FISCAL costing model in an effort to compare “like with like”. However, with the advancements in the computer architecture and infrastructures, it became increasingly difficult to compare performance of various computer systems by merely looking at their specifications. Therefore, domain-specific benchmarking tests (such as NAS Parallel Benchmark and HEPSEPC06) are required and were adopted to compare the systems across HPC, HTC and Cloud infrastructures. In terms of scope, the benchmarking effort focused on the deployment and testing of the equivalent instances, while comparing the HPC and HTC infrastructures against the commercial Cloud offerings (e.g. Amazon EC2). Amazon EC2 is chosen for performance comparisons as it is one of the market leaders and public benchmarking figures are available for the EC2 to compare with. Moreover, it is important to note here that additional performance tests (e.g. MPPTest, STREAM) can be carried out to individually measure various performance metrics (e.g. network performance, memory bandwidth, MPI communication). But, in the context of the e-FISCAL methodology, the aggregated performance factor is sufficient enough to yield the performance adjusted price figures for various infrastructures. The analysis is presented in Appendix 7.7.

The benchmarking exercise also had important secondary goals. First of all, setting up the benchmarking environments within both in-house and Cloud infrastructures gave important tacit information about the practical migration effort that would be necessary in order to move applications from the in-house infrastructures to the Cloud and vice versa. Despite the fact that portability and rapid deployment are the key design goals of benchmarking suites, the effort needed for setting up and maintaining the environments was considerable. This points out towards a need for further studies with real applications when estimating the feasibility of various hybrid e-Infrastructure scenarios. Furthermore, the benchmarking study served an important role in building a common understanding of the e-Infrastructure. Discussing the benchmarking methodologies and their limitations with financial experts had an important role in enabling efficient trans-disciplinary collaboration; in fact this was equally important as discussing the strengths and limitations of the different cost assessment methodologies with e-Infrastructure experts.

4. Results and analysis

This section presents the results of the e-FISCAL project and it is developed along the lines of the research questions raised in the beginning of the document. The analysis is based on 28 questionnaires.

4.1 Respondents characteristics

Out of the 21 countries that had shown interest in participating in the e-FISCAL project (through visiting and browsing through the on-line questionnaire) we have representatives from 16 countries in our final sample[44]. Therefore the set of responses and interest in the survey satisfies one of the target values set at the outset of the project that of having more than 50% of the countries (NGIs or National HPC centres) compared to the invited NGIs ones participating in the study (21 counties out of the 33 countries participating to EGI.eu have shown interest while 16 out of the 33 completed the questionnaire)[45.]

	Country Name	Number of questionnaires
1	Belgium	5
2	Bulgaria	1
3	Cyprus	1
4	Finland	1
5	Germany	1
6	Greece	4
7	Hungary	1
8	Ireland	1
9	Italy	1
10	Latvia	1
11	Norway	1
12	Poland	1
13	Romania	1
14	Spain	6
15	Turkey	1
16	UK	1
	Total	**28**

Table 1 – Countries contributing to e-FISCAL survey

Besides the 28 completed questionnaires reported above there were another 10 questionnaires that were not suitable for processing[46]. In order for a questionnaire to be considered as completed and suitable for further analysis, it should contain information corresponding to more than 75% of the questionnaire categories.

[44] The majority of answers came through the SurveyMonkey software.

[45] The National Grid Initiatives (NGIs) are organisations set up by individual countries to manage the computing resources they provide to the European Grid Infrastructure (EGI). Information about the NGIs is found in http://www.egi.eu/about/ngis/index.html.

[46] Three out these 10 questionnaires corresponded to early attempts of filling in the questionnaire that were eventually completed and therefore included in the sample while the other 7 ended up having excessive missing information and were excluded for the subsequent analysis.

14 out of the 28 respondents provide both computing and coordination activities while 13 provide only computing services. Only one respondent was solely a coordination organisation.
Out of the 28 respondents, 14 (50%) indicated that their institute is only part of NGI/EGI, 3 reported participating into National HPC infrastructure/PRACE, while another 8 participate in both.

The respondents belong to institutes that have multiple roles ending up in a vast number of combinations. The majority of institutes participating in e-FISCAL survey (20 out of 28) are NGI Resource Centres (CPU, storage, etc.), while another 15 out of 28 are HPC centres. As a result, there is no easy separation between HTC and HPC centres, especially when comparing the high-end HTC ones with the low-end HPC ones[47]. Details about the characteristics of the sample are found in Appendix 7.4.

4.2 Research question 1: Analyse the cost structure of HPC and HTC centres

In order to analyse the cost structure of HPC and HTC centres, we used the cost model presented in section 3.2.1. The analytical steps in cost calculations are presented in a comprehensive and transparent way in Appendix 7.3. The descriptive statistics of our sample are presented in detail in Appendix 7.4. The project aimed at calculating the costs for both 2010 and 2011. In this deliverable, we focus on the 2011 cost data as the cost data of 2010 seemed less relevant at the time the report was written (January 2013)[48]. However, it should be noted that costs between 2010 and 2011 are decreasing in-line with the trend of lower hardware prices and better overall efficiency. In order to get a multifaceted view of the cost structure of HPC and HTC centres, we treat input data under three different option cases.

Option 1: Basic Case: We run the e-FISCAL costing model by using as inputs the average and the median[49] values of all necessary input data from the total e-FISCAL sample (i.e. the average and the median value of acquisition cost for CPUs for the total sample, the average and median depreciation rates of the total sample etc.). Even though this procedure may not adequately account for the existence of economies of scale and does not correspond to the specific circumstances of each and every participant in the study, it makes a fair approximation of the "average" case. Additionally, the core unit of analysis permits the presentation of results in a way that is easy to understand and follow. It is a therefore a reference point that could be used for high-level comparisons.

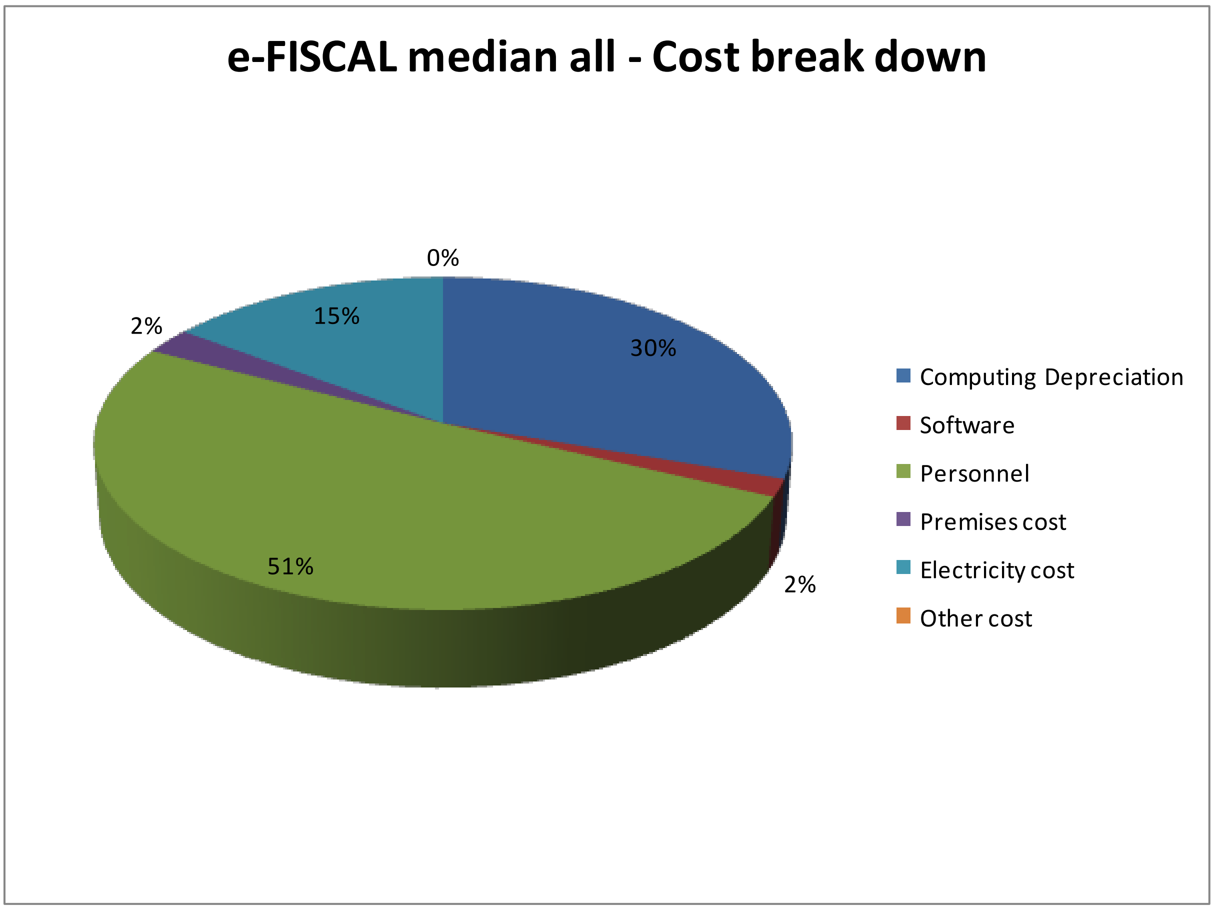
The results are presented as follows (detailed calculations are found in Appendix 7.3):

[47] This was a comment raised by the reviewers of one of the intended publications of the project (for the eChallenges 2012 conference) and was also raised with the e-Infrastructure community during the e-IRG workshop. As there are several definitions of these two terms, some of which are not satisfactory (as the Wikipedia one), the project decided to raise the attention of the e-Infrastructure community and especially EGI and PRACE. In the Terms and Definitions section we have adopted the EGI definition.
[48] A thorough analysis of the 2010 cost information and break down is found in D2.2: Computing e-Infrastructure cost calculation at national and European level found in http://www.efiscal.eu/deliverables.
[49] The median is described as the numeric value separating the higher half of a sample, a population, or a probability distribution, from the lower half.

Amounts in €	e-FISCAL average all	e-FISCAL median all
Total yearly CAPEX/core	109	63
Total yearly OPEX/core	307	145
Total yearly cost/core	416	208
Utilisation rate	65%	75%
Cost per core hour	0,0730	0,0317
CAPEX as % of total costs	26%	30%
OPEX as % of total costs	74%	70%
FTEs/1,000 cores	4,61	2,16

Table 2: Cost information for Basic Case– all respondents

In the Graph below (Graph 1) the cost breakdown of the total e-FISCAL sample in main cost categories is presented.



Graph 1: e-FISCAL total sample cost break down (median)

From the information presented above the following observations can be made:
The e-FISCAL median cost per core hour (which takes into account the so-called "outliers" that are either very high or low compared to typical values) is around €0.032 per core hour, while the e-FISCAL average cost per core hour is around €0.073 core hour. The breakdown between CAPEX and OPEX is around 30%-70% (median) to 26%-74% (average). The average utilisation rate used to calculate the average and median cost per core hour for 2011 is 65% and 75% respectively. This refers to a mixture of HTC and HPC sites. As an example, for 2011, EGI reported a

utilisation rate of 71.3%. The utilisation rate in the e-FISCAL project has been calculated by taking into account yearly logical wall clock time and available CPU cores at the end of the period (i.e. 31/12/2011). This assumes that the number of CPU cores reported at the end of the period is available throughout the year. If this is not the case (e.g. because of an infrastructure upgrade towards the end of the year), the utilisation rate calculated is underestimated.

Other interesting findings are the high numbers of depreciation rates for the hardware (average 5 years), the quite good rates of PUE (of around 1.5 median value) and the percentage of electricity cost (around 15% median value of all costs).

51% of total costs (median values) is dedicated to personnel costs. It is expected that the personnel costs for a very large computing centre (in the order of 100,000 cores) can show economy of scale if compared to the same capacity distributed in smaller sites that are federated together. We discuss this issue in more depth later on. However, the fact that staffing represents a very significant cost category was expected. Intersect 360 in both its reports of 2011 and 2012 (Intersect 360, 2011; 2012) mentions that the staffing represents 23% of top–level budgets. This amount in not directly comparable to e-FISCAL due to the different methodology employed (Intersect 360 conducts a census on budgets and as such the total acquisition budget of hardware-and not its depreciation- is included in the 100% calculation); however, the magnitude is similar. Moreover, they claim that they expect to see in the near future an increase on the money spent on people to manage and maintain systems, as the computers and overall IT architectures become more complex.

Finally, our results seem to coincide with earlier findings found in literature as shown in Table 3. However, while other studies concentrate only on site or centre, e-FISCAL is the first costing study that comes up with conclusions resulting from an extensive synthesis.

Reference	Cost per core hour	Comments
Hawtin et al. (2012)	€0.075	Study for JISC UK - Differences between institutions reviewed
US DoE - Magellan report (2011)	€0.015	Hopper system – National Energy Research Scientific Computing Centre- including storage sub- system
Smith (2011)	€0.031	Purdue campus, USA
University of Washington	€0.020	Hyak cluster, USA
Cohen and Karagiannis (2011)	€ 0.09 – € 0.14	Stratified sample of EGI centres - Assuming 60% utilization ratio – storage cost included, (**costs refer to 2009**)
Cohen and Karagiannis (2011)	€0.08 – €0.10	Stratified sample of EGI centres - Assuming 60% utilization ratio – storage cost excluded (**costs refer to 2009**)

Table 3 – Literature review summary in relation to Cost per core hour or Cost per core hour

Option 2: Basic Case Split: We also run the e-FISCAL costing model by using as inputs the average and the median values of all necessary input data from the HPC and the HTC e-FISCAL sub-samples (e.g. the average and the

median value of acquisition cost for CPUs for all HPC centres, the average and median depreciation rates of all HTC centres, etc.). Therefore, we split the total e-FISCAL sample in HPC and HTC subsamples. It has to be noted that the classification of centres into HPC and HTC is based on assessments by respondents. Some centres are classified as both HPC and HTC. Again, the logical CPU unit of analysis permits the presentation of results in a way that is easy to understand and follow and constitutes a benchmark that could be used for high-level comparisons.
The results are as follows:

Amounts in €	**e-FISCAL HPC average**	**e-FISCAL HPC median**
Total yearly CAPEX/core	84	58
Total yearly OPEX/core	217	124
Total yearly cost/core	301	182
Utilisation rate	66%	75%
Cost per core hour	0,0520	0,0277
CAPEX as % of total costs	28%	32%
OPEX as % of total costs	72%	68%
FTEs/1,000 cores	2,90	1,59

Table 4: Cost information for Basic Case Split – HPC centres

Amounts in €	**e-FISCAL HTC average**	**e-FISCAL HTC median**
Total yearly CAPEX/core	85	59
Total yearly OPEX/core	325	169
Total yearly cost/core	411	229
Utilisation rate	59%	74%
Cost per core hour	0,0795	0,0353
CAPEX as % of total costs	21%	26%
OPEX as % of total costs	79%	74%
FTEs/1,000 cores	5,86	3,69

Table 5: Cost information for Basic Case Split – HTC centres

Graphical representation of cost breakdown is found in Appendix 7.3.

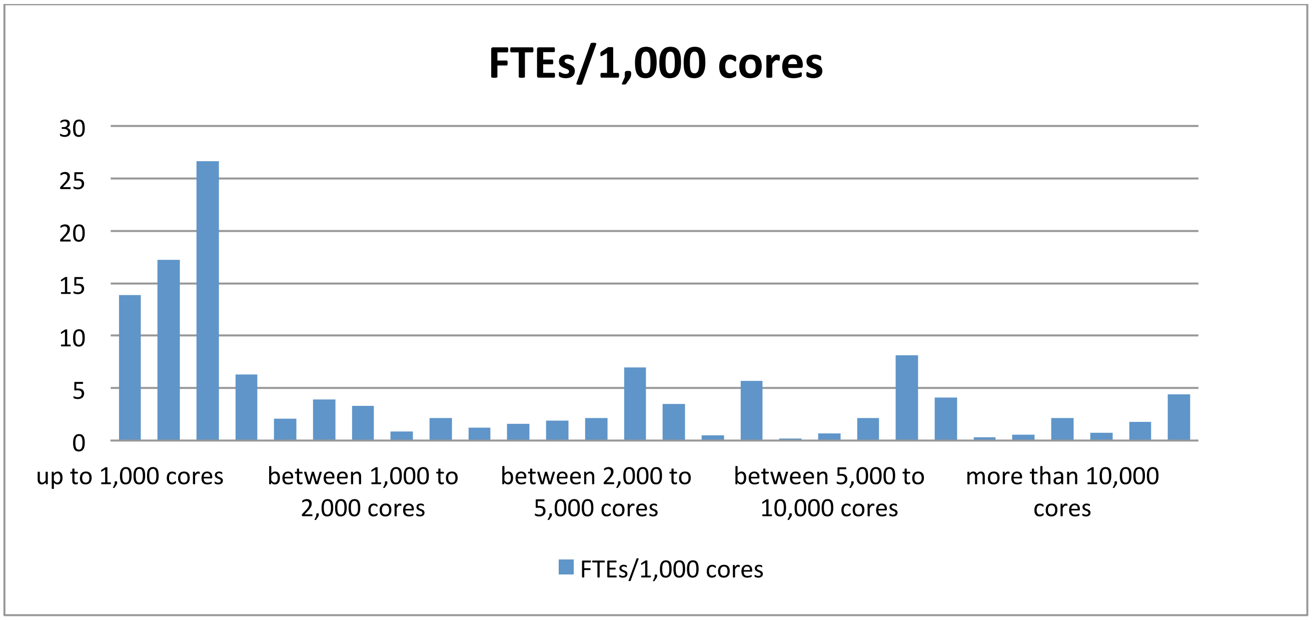
Option 3: Case-by-case analysis: Finally, we run the e-FISCAL costing model on a respondent by respondent basis (we call this case–by–case analysis). Therefore, we used as inputs the data submitted by each respondent for every cost category. In case of missing cost information we proceeded in replacing the missing information with the average value of the input parameter corresponding to the e-FISCAL total sample. This analysis is also used as a cross check to the robustness to the "basic case" presented above. We present median values as, as they are not influenced by outliers, they provide more representative indications of the "usual" case.

The results are as follows:

Amounts in €	**e-FISCAL Case by case all - Median**	**e-FISCAL case by case HPC - Median**	**e-FISCAL case by case HTC Median**
Cost per core/year	358.74	258.45	390.81
Cost per core hour	0.058	0.039	0.066

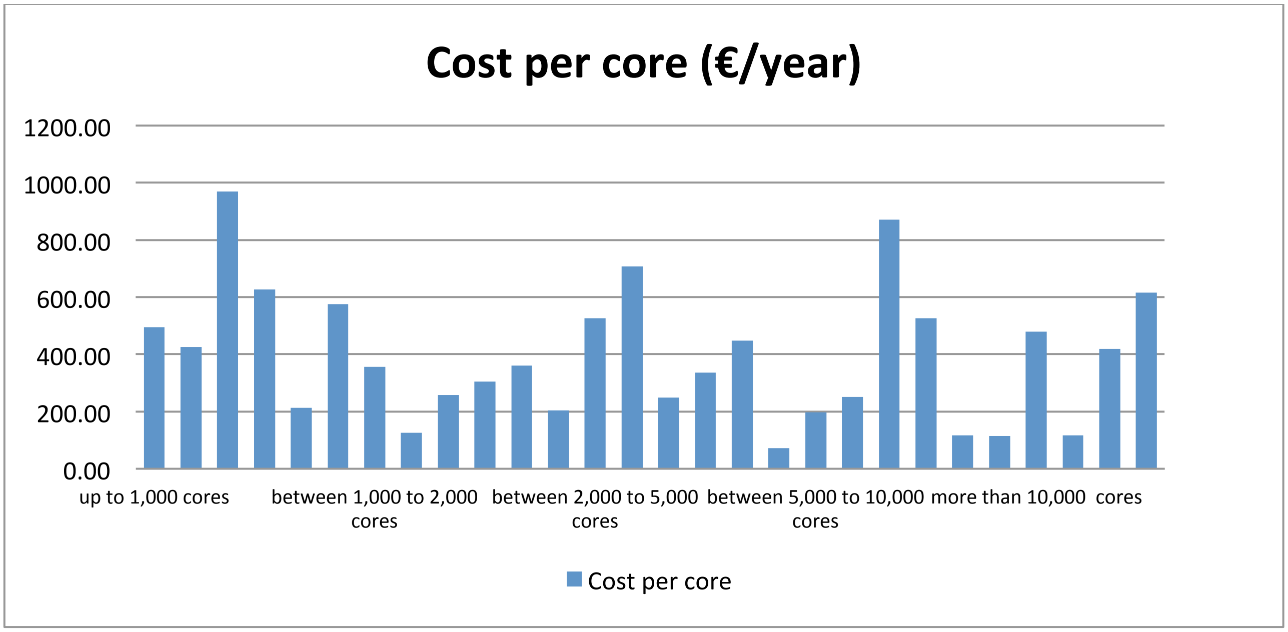
Table 6: Cost information for Case by Case Analysis/ all centres- HPCs - HTCs

The analysis above permits the identification of ranges within the different cost parameters seem to lie. In all cases, the median cost per core hour and the median cost per core calculated on the basis of case–by-case analysis lies within the boundaries identified by the corresponding basic case.
The case-by-case analysis also permits the exclusion of some interesting analyses that relates: a) to the cost per core hour on the basis of the site size, b) the number of FTEs/1,000 cores on the basis of the site size and c) the cost breakdown per site size.
The graphs below do not intend to be precise, but to provide an overview of cost aspects relating to size. The presentation that follows also serves the preservation of the anonymity of the sites contributing to e-FISCAL survey.



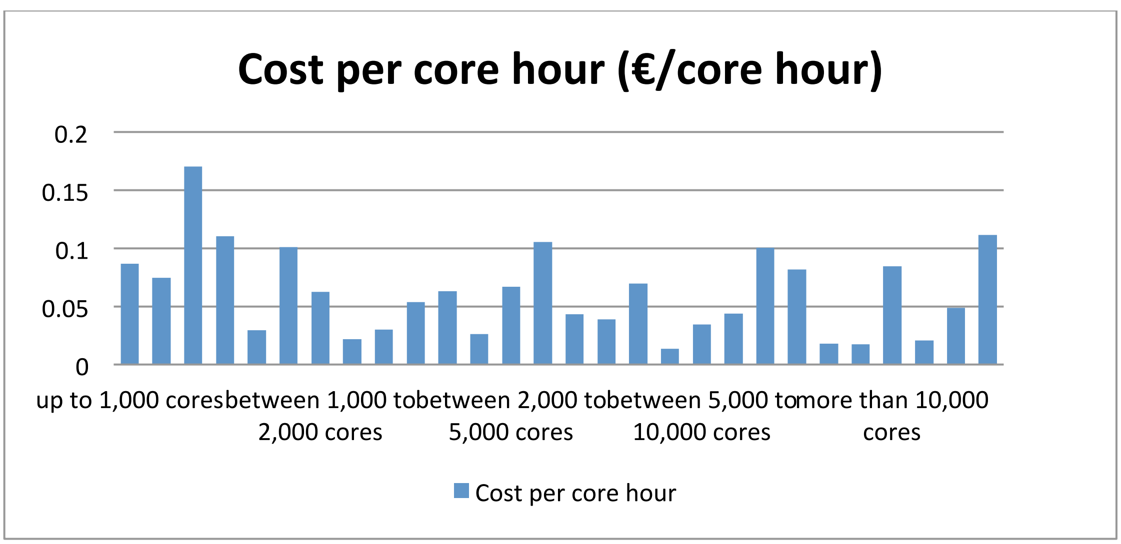
Graph 2: Number of FTEs per 1,000 cores

From Graph 2, it is evident that the ratio of FTEs/1,000 CPUs is better for the larger sites compared to the smaller ones. The same conclusion is derived by using the spearman correlation. The correlation between number of cores and FTEs/1,000 cores has a statistically significant relation of -0.4111 at 5% statistical significance level. The person correlation reveals again statistically significant relation of -0.3354 albeit at 10% statistical significance level. The larger the site the lower the number of FTEs/1,000 cores.



Graph 3: Cost per Core

From Graph 3 where the cost per core/year is plotted against the size of sites, no unanimous conclusions can be derived. The cost per core/year is influenced by several parameters such as technology, number of FTEs, efficiency in energy consumption, etc. However, these results reflect that significant economies of scale are not evident at least on the basis of the input we had processed in the questionnaires. The existence of economies of scale would be translated in the cost per core for larger sites to be less than smaller sites, ceteris paribus, due to the division of fixed costs to more CPU cores.



Graph 4: Cost per Core hour

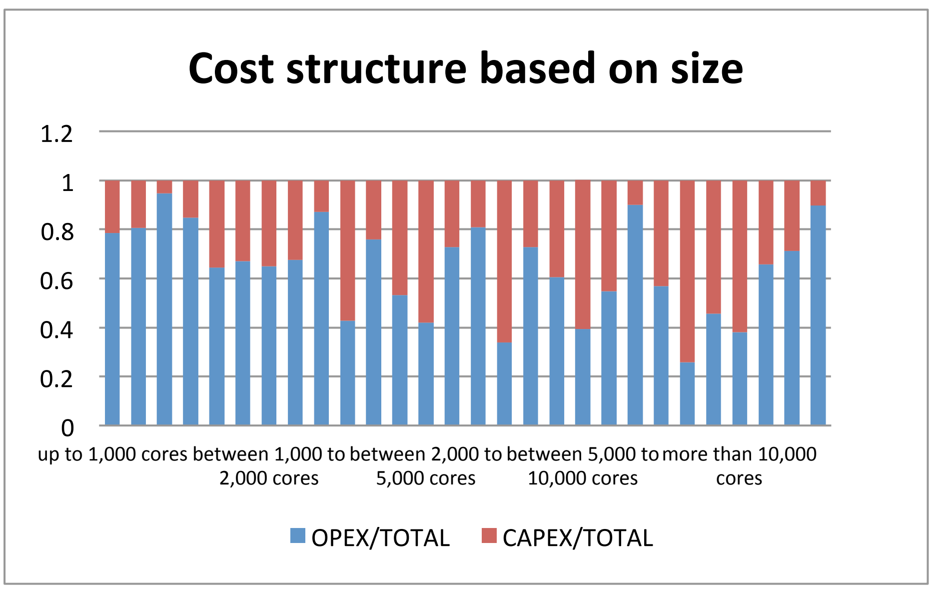
Finally, as shown in Graph 4, the cost per core hour is higher in the very small sites up to 1,000 cores. For the rest of the sample there is no clear trend that can be easily identified. However, the lower costs per core hour are evidenced in larger, in terms of number of cores, sites. More specifically, it appears that sites with <1000 cores are much more expensive than sites >1000 cores. Indicative (median and average) values for cost per core hour are around € 0.09 for the first category and € 0.05 for the latter.

In order to assess whether there is a statistically significant relationship between the size of the sites and the number of FTEs, the cost per core and the cost per core hour for 2011, we have also run univariate Ordinary Least Square Regressions. The results are shown in Appendix 7.5 and confirm that a linear relationship between cost aspects and the size of a site in terms of CPU cores cannot be supported. However, there is a statistically significant negative relation between the number FTEs/1,000 cores and the size of the sites in cores. Therefore, OLS confirms the correlation results.

As a final comment, it must be noted that there are parameters that may affect the cost hour cost among respondents. A summary of these factors follows:

- **Utilisation rates**: The utilisation rate (number of CPU hours actually used compared to the theoretical maximum) plays an important role in calculating the cost per core/hour. Core hours that are not used could be considered as are wasted resources whose cost is added to the hours been used (Hawtin et al., 2012).
- **Depreciation rates**: The prolongation of hardware and auxiliary equipment useful life ends up in decreasing the yearly cost per core CPU and the corresponding cost per core hour.
- **Salaries and premises costs:** These costs have a very local focus and are highly influenced by national circumstances.
- **Electricity effectiveness and cooling:** The PUE of the respondents as well as the climate conditions that prevail in the site area affect electricity consumption which in turn affects OPEX calculation.

Finally, we present the cost breakdown into CAPEX and OPEX in all our respondents.



Graph 5: CAPEX/OPEX per site based on size

The graph (Graph 5) is indicative of the fact that there is heterogeneity in the cost structure of sites/centres. However, it is rather evident that in the vast majority of cases OPEX is not less than 40% of total costs.

4.3 Research question 2: Approximate the total yearly cost of HPC and HTC in Europe

In order to approximate the total yearly cost of dedicated HPC and HTC e-Infrastructures we use the number of cores available as of 2011 and we multiply with the average and median cost per core per year under Option 2 and Option 3 presented above.

We used the number of cores of EGI as of1[50] in order to calculate the cost of European dedicated HTC centres.

Number of cores as of 2011	306,684	Amounts in '000 €
Option 2: Yearly Cost per core (average 2011) - HTC	€ 411	€ 126,047
Option 2: Yearly Cost per core (median 2011) - HTC	€ 229	**€ 70,230**
Option 3: Yearly Cost per core (median 2011) - HTC	€ 390.81	**€ 119,855**

Table 7: Approximation of HTC in Europe

Therefore the total yearly cost of HTC has been assessed in the range of €70,000 K to €125,000 K.

In order to approximate the total yearly cost of the European dedicated HPC e-Infrastructure by taking into account the cost per core per year of HPC centres. We have to note that in our sample there were not Tier-0 sites. Therefore our metrics may not be suitable for approximating the cost of Tier-0s[51].

In order to approximate the cost of the dedicated European HPC e-Infrastructure, we used the number of cores from PRACE Tier-0 centres added together with the number of cores from other national Tier-1 centres[52].

Number of cores as of 2011	566,868	Amounts in '000 €
Option 2: Yearly Cost per core (average 2011) - HPC	301	**170,627**
Option 2: Yearly Cost per core (median 2011) - HPC	182	103,170
Option 3: Yearly Cost per core (median 2011) - HPC	258.45	146,507

Table 8: Approximation of HPC in Europe

The yearly cost of the dedicated HPC e-Infrastructure has been assessed in the range of €100,00K to €170,000K.

[50] https://www.egi.eu/infrastructure/operations/figures_and_utilisation/index.html

[51] However, we crosschecked our analysis results with publicly available information found in http://spscicomp.org/wordpress/wp-content/uploads/2011/05/brehm-LRZ-and-SUperMUC_small2.pdf

[52] http://www.prace-ri.eu/IMG/pdf/d9.2.1_1ip.pdf

4.4 Research question 3: Cost of HPC and HTC centres compared to commercial Cloud Offerings

The pricing policy of a company constitutes a tool in order to fulfil its market and strategic goals. Price is definitely not synonymous to value and it is not necessarily cost based[53]. Therefore the only reason why we compare costs with prices is in order to assess whether the cost of resources consumed for in-house computing is above or lower compared to commercial cloud prices. The outcome of this comparison is expected to provide indirect indications on the efficiency of in-house computing, under two major assumptions: a) that cloud prices are based on costs plus a mark-up and b) that clouds services are the “same” with in-house services. While the first hypothesis cannot be tested as cloud service providers do not reveal their cost structure and profit margins per service category, we try to challenge the second through the incorporation of the benchmarking results (see section 4.4.2). Another important dimension is that prices usually change more easily than costs in the short run and this change in prices does not necessarily mean that there have been cost savings achieved. Notwithstanding the above limitations, having a better insight in the relation between the cost for in-house service provision and commercial prices would provide useful information for decision making. It should however be noted that this comparison does not reveal in any way the savings a site could achieve by using cloud services. This is a totally different exercise that should concentrate on avoidable costs and is out the scope of the project.

In order to compare the costs calculated in e-FISCAL with the prices of commercial cloud offerings we had to make a selection from a plethora of providers. The prices of commercial cloud offerings in the market vary not only in monetary terms (i.e. prices charged) but also on the characteristics of the services offered (e.g. mainly technical characteristics and configuration of hardware). We discuss pricing policies of commercial cloud providers in section 4.5.2. There are already sites that help potential customers to choose between alternative providers by identifying the prices charged for a chosen service[54].

For the purpose of our comparison we have chosen Amazon that provides Elastic Compute Cloud (Amazon EC2)[55] and OVH[56]. The selection is made on the basis of the following criteria. Amazon is considered to be one of the world’s largest cloud computing operations. Amazon’s prices have been used in several comparative studies (Magellan report, 2011; Hawtin et al., 2012). Moreover, Amazon has a rather complex but rather transparent pricing system (provides free trial of services, on demand instances, reserved instances[57], spot instances and volume discounts among others)[58]. It is a US based company.

---

[53] For more information about these distinctions please go to the material presented in the final e-FISCAL workshop in Amsterdam http://www.efiscal.eu/files/presentations/amsterdam/E-FISCAL-Tutorial_Vfinal.pdf.

[54] Examples of such sites are: http://www.cloudorado.com, http://cloudpricecalculator.com/, http://cloud-computing.findthebest.com/, http://www.cloudsurfing.com/ (sites accessed on 31/1/2013).

[55] http://aws.amazon.com/ec2/

[56] http://www.ovh.co.uk/cloud/

[57] With reserved instances, you pay a fixed up-front cost in order to pay a lower per hour cost. If the instance is used for a majority of the reserved period (one year in our analysis), this results in a lower effective rate.

[58] The analytical pricing information of Amazon is found in http://aws.amazon.com/ec2/pricing/

OVH is a privately owned company in France that provides dedicated servers, mutual hosting, domain names and VOIP telephony services. The company offers localized services in France, Germany, Italy, Poland, Spain, Ireland, United Kingdom, the Netherlands and Finland[59]. OVH has reported that it has reached 120,000 physical servers hosted in its datacenters[60]. The pricing policy of OVH for cloud services is simple, transparent and rather straightforward. OVH is a Europe-based company. While the costs of in-house computing are on core hour basis the price charged by commercial providers is per instance hour. Therefore we had to assess the number of cores per instance and convert the commercial prices into price per core hour. As for Amazon, the type of instances that are relevant for e-FISCAL (http://aws.amazon.com/ec2/instance-types/) are the M/L/XL standard instances and the cluster compute quadruple extra-large instance[61].

Therefore the on-demand prices for these services as well as the price of the reserved instances were used. Spot instances[62] are not considered relevant as Amazon proposes that spot Instances are better suited for time-flexible, interruption-tolerant tasks and warrens the user to be always prepared for the possibility that his/her spot instance is interrupted. Magellan (2011) does not also use prices of spot instances for comparisons. They support their decision on the fact that the runtime for a spot instance is unpredictable and application programmers need to design their applications to handle pre-emption, which would not match the requirements for their applications. However, spot pricing does provide an estimate of the absolute lowest bound for pricing, since it essentially reflects the price threshold at which Amazon is unwilling to offer a service.
The cost per core hour for the EC2 services was calculated as followed:

1. The prices for on demand instances and reserved instances were identified in Amazon's site (http://aws.amazon.com/ec2/pricing/) for both Linux and Windows. The prices refer to EU (Ireland) but when such services are not provided in the EU the prices of West Virginia are used for purposes of completeness of the analysis. All reserved instances refer to heavy use.

---

[59] http://en.wikipedia.org/wiki/OVH#cite_note-dcn2009-12
[60] http://www.datacenterknowledge.com/archives/2009/05/14/whos-got-the-most-web-servers/ (accessed 31/1/2013)

[61] In order to compare the like with like, a mapping has been done between the HPC/HTC infrastructures specified by the respondents of the e-FISCAL survey and the Amazon EC2 instances. Below are the two working examples of how this mapping is achieved. Case 1: Centre X has specified its infrastructure as follows: 2 hexa-core Xeon 5650 "westmere" 2.66GHz cpu's 24GB RAM. Now, each compute node is equivalent to 12 (2x6) cores and it can be roughly mapped to 1.5 X Cluster Compute Quadruple X-Large instance from EC2. The total number of cores in a Cluster Computer Quadruple X-Large is: 2 X 4 = 8, so 1.5 X 8 = 12 cores and it has 23 GB memory. Case 2: A Centre Y has specified its infrastructure as follows: HP Bl2x220c G5, L5420, 4 core/CPU, 2.5 GHz, 16 GB RAM. Now, each compute node is quad core and it can be mapped to M1 X-Large instance within EC2 infrastructure. M1 X-Large instance has 4 virtual cores and it has 15 GB of memory.
[62] Spot Instances allow the user to name his/her own price for Amazon EC2 computing capacity. The user bids on spare Amazon EC2 instances and run them whenever the bid exceeds the current Spot Price, which varies in real-time based on supply and demand. According to Amazon the Spot Instance pricing model complements the On-Demand and Reserved Instance pricing models, providing potentially the most cost-effective option for obtaining compute capacity, depending on the application. While spot instances perform exactly like other Amazon EC2 instances while running they differ, except for lower price, in the fact that there is a possibility of being interrupted. This happens when the Spot price exceeds the max bid.

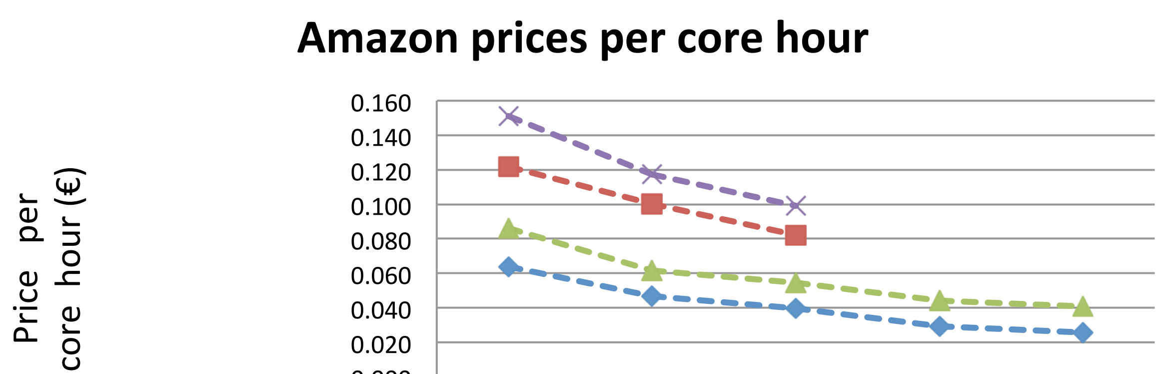
2. The instance cost per hour for the reserved instances for one and three years was calculated by taking into account a 60% and 80% usage rate. These rates are within the range of the utilization of the in-house infrastructure of the sample sites.
3. The cost per instance hour was then transformed into cost per core hour by taking into account the number of cores an instance could refer to.

The equivalence of the EC2 instances with the number of cores in present in the following Table (Table 9).

Amazon EC2	
Instance type	**Number of cores**
M1 Medium	2
M1 Large	4
M1 Extra Large	8
Cluster Compute Quadruple Extra Large	8

Table 9: Number of cores for transforming Instances to number of cores

The Amazon prices transformed on a per core hour basis are therefore as follows:



	On demand	Cost per core hour (€/hour) - 60% - 1year reserved	Cost per core hour (€/hour) - 80% - 1year reserved	Cost per core hour (€/hour) - 60% - 3-years reserved	Cost per core hour (€/hour) - 80% - 3 years reserved
Standard M/L/XL - Linux	0.064	0.047	0.040	0.029	0.026
Cluster Compute Quadruple Extra Large *- Linux	0.122	0.100	0.082		
Standard M/L/XL - Windows	0.086	0.062	0.055	0.044	0.041
Cluster Compute Quadruple Extra Large *- Windows	0.151	0.117	0.099		

Notes:

Prices have been transformed from $ with exchange rate 1€ =$ 1.33270, http://www.alpha.gr/tools/eeurocur.htm (accessed on 16/1/2013). Amazon prices as on 16/1/2013. The prices refer to EU (Ireland). The prices of cluster compute quadruple extra-large are not provided in EU; West Virginia prices were used. No price means no availability of the service.

Graph 6: Amazon prices transformed on €/core hour

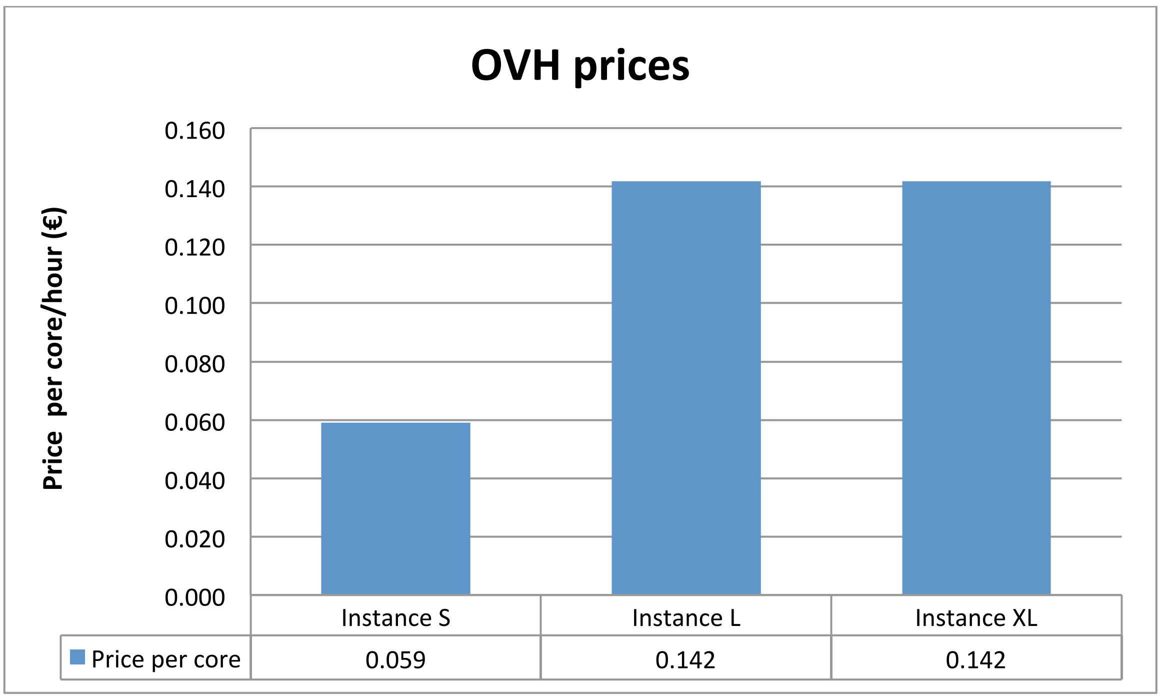
It has to be noted that while standard M/L/XL instances have different prices per instance hour, the cost per core hour is the same.

The prices per hour of the OVH instance offers[63] were also translated into price per core hour as follows:

OVH	
Instance type	**Number of cores**
S	1
L	2
XL	4

Table 10: Number of cores for transforming Instances to number of cores

The OVH prices on per core hour basis are therefore as follows:



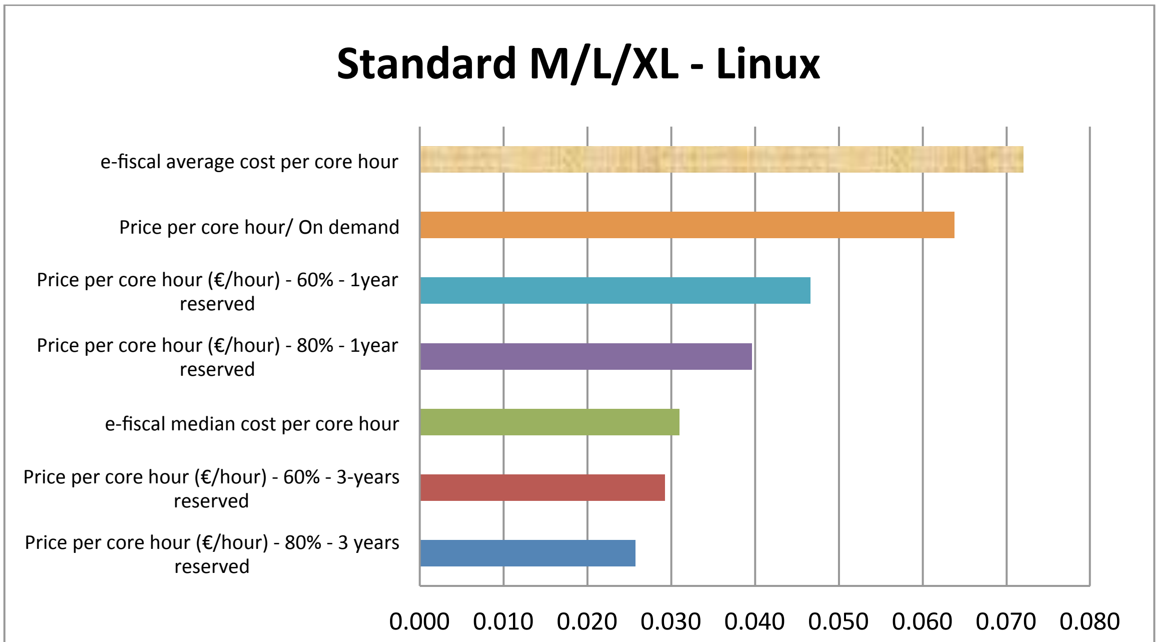
Notes:
Prices have been transformed from £ to € with the exchange rate 1€ = £ 0.8464, http://www.alpha.gr/tools/eeurocur.htm (accessed on 31/1/2013). OVH prices as of 31/1/2013.

Graph 7: OVH prices transformed on €/core hour

63 http://www.ovh.co.uk/cloud/instances/index.xml#instances

4.4.1 No performance adjusted comparison – Amazon

In the following table the comparison between the cost per core hour for 2011 streaming from e-FISCAL calculations **(Basic Case)** and the Amazon prices for Standard M/L/XL instances - Linux per core hour are presented. We concentrate our discussion on Linux as it more commonly used in research. Nevertheless, we also present the Windows comparisons in the Appendix 7.6.

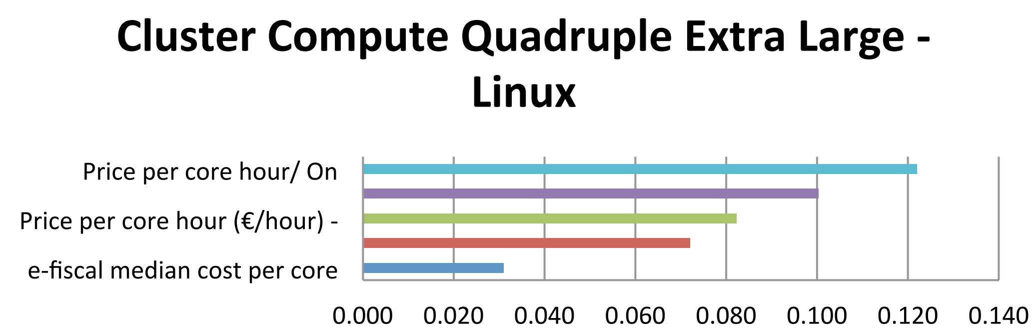


	Price per core hour (€/hour) - 80% - 3 years reserved	Price per core hour (€/hour) - 60% - 3-years reserved	e-fiscal median cost per core hour	Price per core hour (€/hour) - 80% - 1year reserved	Price per core hour (€/hour) - 60% - 1year reserved	Price per core hour/ On demand	e-fiscal average cost per core hour
Standard M/L/XL - Linux	0.026	0.029	0.031	0.040	0.047	0.064	0.072

Graph 8: Comparing e-FISCAL cost with Amazon prices (Standard M/L/XL instances – Linux)

It is evident that the e-FISCAL median cost (basic case) is more expensive than the prices of the 3-year reserved instances but less expensive than all other cases. The average e-FISCAL cost (basic case) is more expensive than even the on-demand instances.

The comparison between the cost per core hour for 2011 streaming from e-FISCAL calculations (Basic Case) and the Amazon prices for Cluster Compute Quadruple Extra Large - Linux per core hour is presented below.

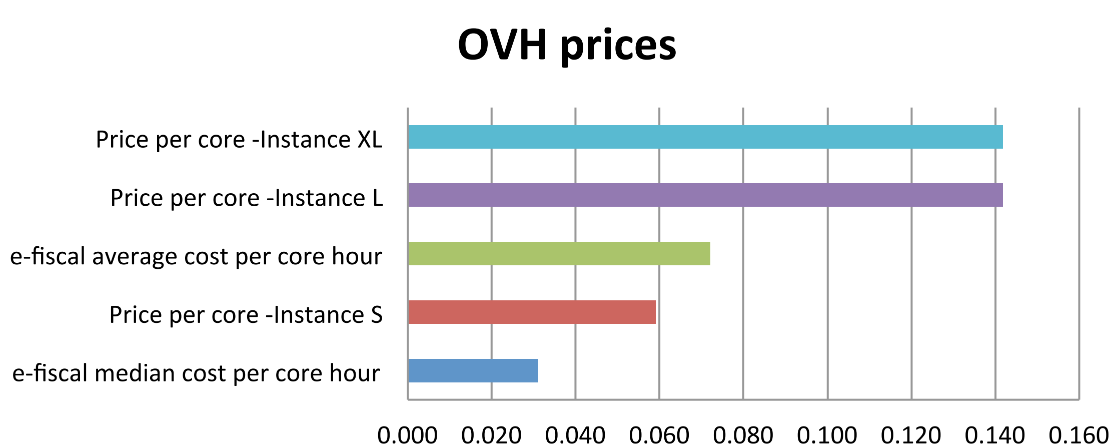


	e-fiscal median cost per core hour	e-fiscal average cost per core hour	Price per core hour (€/hour) - 80% - 1year reserved	Price per core hour (€/hour) - 60% - 1year reserved	Price per core hour/ On demand
Cluster Compute Quadruple Extra Large - Linux	0.031	0.072	0.082	0.100	0.122

Graph 9: Comparing e-FISCAL cost with Amazon prices (Cluster Compute Quadruple XL instances – Linux)

It is evident that both the median e-FISCAL cost and the average e-FISCAL cost (basic case) per core hour are less expensive than the prices of both on demand and one-year reserved instances.

As for OVH, the comparison showed that the OVH prices per core hour except for the S instance are higher than the average cost core hour calculated for the total e-FISCAL sample. However, the median e-FISCAL cost per core hour is lower that all three analysed services offered by OVH.



	e-fiscal median cost per core hour	Price per core - Instance S	e-fiscal average cost per core hour	Price per core - Instance L	Price per core - Instance XL
OVH prices	0.031	0.059	0.072	0.142	0.142

Graph 10: Comparing e-FISCAL cost with OVH prices

4.4.2 Performance adjusted comparison (based on benchmarking) – Amazon

We have also compared the prices of Amazon with the costs of the **Basic case** of e-FISCSAL after incorporating in the analysis the results of the benchmarking exercise for both HTC and HPC (details about the benchmarking exercise are found in Appendix 7.7).

The main findings of the benchmarking exercise as well as the way they had been incorporated in the price comparisons are discussed below

HTC vs. Cloud benchmarking

The HTC vs. Cloud benchmarking effort focused on the comparison of the relevant Amazon EC2 instance types versus HTC instances. It is worth clarifying here that the classification of the EC2 instances as an HPC or HTC instance is done based on the Amazon specification for various instances types. The following Table 11 summarises the results of the benchmarks with various different instance and processor types:

Benchmark & Instance Type	Average EC2 Performance impact in %	Ranges for the EC2 performance impact in %
HEPSPEC06 + M1 Medium instance + Sandy Bridge	-48	-3 to -58
HEPSPEC06 + M1 Medium instance + Opteron	-18	-4 to -48
HEPSPEC06 + M1 Large instance + Sandy Bridge	-59	-9 to -57
HEPSPEC06 + M1 Large instance + Opteron	-31	-9 to -49
HEPSPEC06 + M1 X-Large instance + Sandy Bridge	-46	-8 to -56
HEPSPEC06 + M3 X-Large instance + Sandy Bridge	-43	-18 to -55
HEPSPEC06 + M3 Double X-Large instance + Sandy Bridge	-38	-9 to - 55

Table 11: EC2 M1 (M, L, XL) and M3 (XL, Double XL) VS. HTC instances (Intel+AMD) @ INFN

The HEPSEPC06 benchmark is used to compare the Cloud vs. HTC infrastructure. In order to compare the effect of distinct characteristics (i.e. virtualization, multi-tenancy, resource over-subscription) and possible variations within the Cloud infrastructure, HTC instances and benchmarking environment were configured in multiple settings and HEPSPEC06 was executed on each setting. The idea for each individual setting was to obtain the results for performance variations while taking into account the best-case and worst-case scenarios. The best-case scenario here could reflect the situation with many of the existing HTC centres (i.e. no virtualisation or bare-metal) or virtualization with minimum possible workload. On the contrary, the worst-case would include the fully loaded machine with "n" number of virtual machines and all running HEPSPEC06 or even the case when the available resources (e.g. memory) are over-subscribed among the virtual machines.

In order to calculate the performance adjusted prices for comparing the in-house HTC computing costs and Cloud offerings these performance variations (for both best and worst case) are used.

HPC vs. Cloud benchmarking

The results of the benchmarking exercise that refer to the comparison of the in-house HPC instances (i.e. Stokes HPC system) against the Amazon EC2 HPC instances were included in D2.2. The following Table 12 provides a comparison of the benchmark results for both in-house and EC2 HPC instances using the NAS Parallel Benchmark (NPB) MPI and OMP variants.

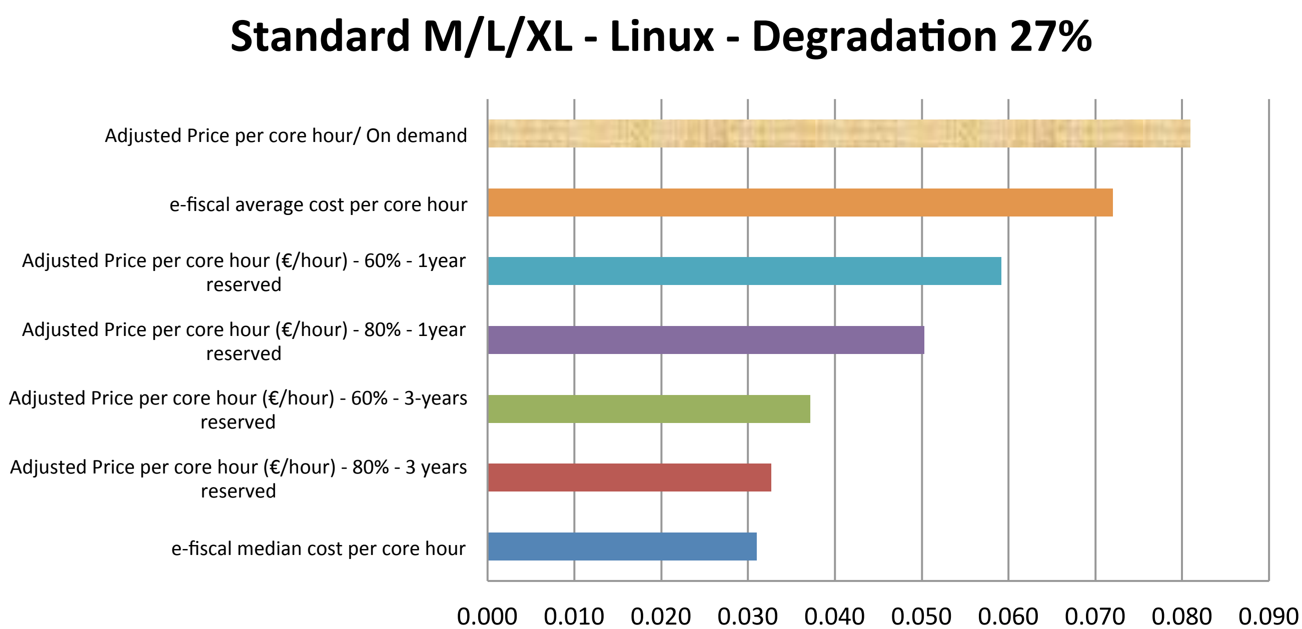
Benchmark	EC2 average performance degradation VS. HPC in %	EC2 range of performance variations VS. HPC in %
NPB MPI – Class B	- 48.42	- 1.02 to - 67.76
NPB OMP – Class B	- 37.26	- 16.18 to -58.93

Table 12: EC2 HPC VS. Stokes HPC instance @ ICHEC

The MPI and OMP variants of NPB are chosen to address the most common parallelism techniques used within the HPC community. Although, medium sized problem (i.e. Class B) was chosen for various NPB kernels (i.e. BT, CG, EP, FT, IS, LU, MG and SP), the individual kernels significantly vary in size and type to cover various aspects of a real-world parallel/HPC application. One may argue that Class B would not be a real representative in terms of scale for a typical HPC/parallel application. But, the fact of the matter is that our objective and scope (as highlighted in the methodology section) is to identify the performance variations across EC2 and in-house HPC infrastructures. Of course, benchmarking a typical large-scale parallel/HPC application would yield more accurate results but would also require significant resources. Nevertheless, the performance variations (in Table 12) are quite evident even with Class B problem. In order to calculate the performance adjusted prices for comparing the in-house HTC computing costs and Cloud offerings these performance variations are used.

Therefore, the outcomes of the benchmarking exercise were used in an attempt to compare like with like services. More specifically, the loss in performance has been embedded in the calculations as follows: Price x (1+ % loss in performance). For example, the price for the on demand instances per core hour for M/L/XL standard instances is €0.064/ core hour. By taking into account a performance degradation of 27% (as evidenced in the benchmarking exercise) the performance-adjusted price goes to € 0.081 [(0.064 + (1+0.27)) = 0.081]

By comparing the cost per core hour for 2011 streaming from e-FISCAL calculations **(Basic Case)** and the Amazon **performance adjusted prices** for Standard M/L/XL instances - Linux per core hour we get the following results. The performance degradation used to adjust the prices of Standard M/L/XL instances is 27%

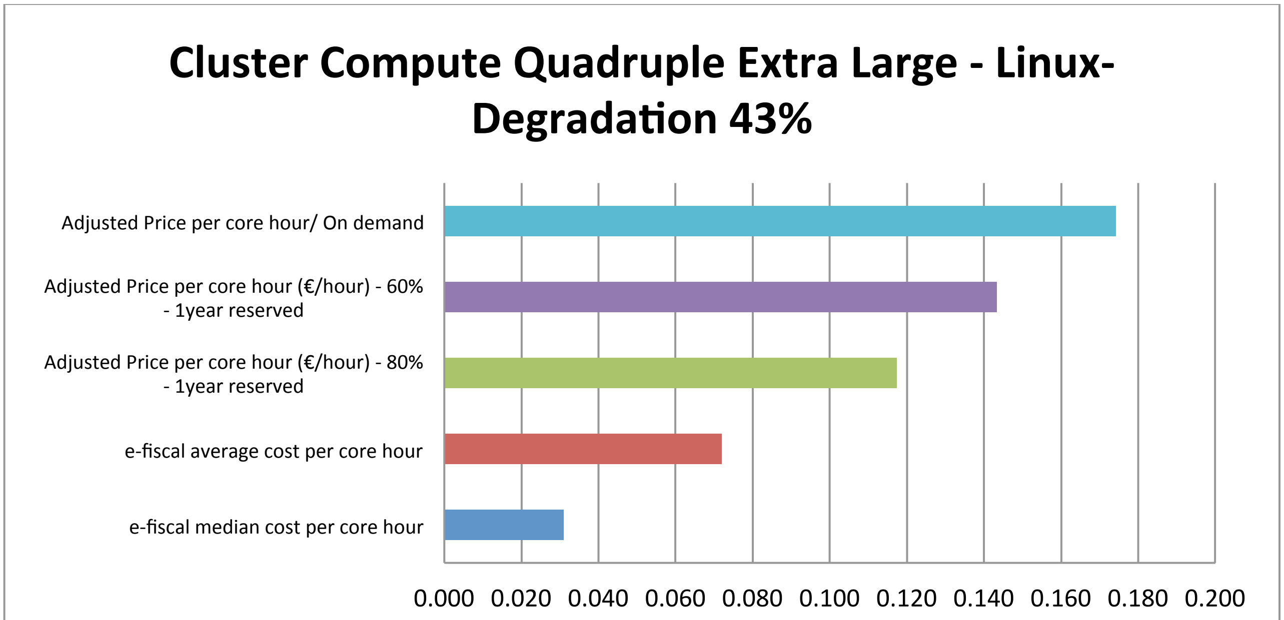


	e-fiscal median cost per core hour	Adjusted Price per core hour (€/hour) - 80% - 3 years reserved	Adjusted Price per core hour (€/hour) - 60% - 3-years reserved	Adjusted Price per core hour (€/hour) - 80% - 1year reserved	Adjusted Price per core hour (€/hour) - 60% - 1year reserved	e-fiscal average cost per core hour	Adjusted Price per core hour/ On demand
Standard M/L/XL - Linux	0.031	0.033	0.037	0.050	0.059	0.072	0.081

Graph 11: Comparing e-FISCAL cost with Amazon adjusted for performance prices (Standard M/L/XL instances – Linux)

By taking into account the performance adjusted prices the median e-FISCAL cost per core hour (basic case) is less expensive even than the 3-year reserved instances. Therefore it is less expensive in all cases.
The average e-FISCAL cost per core hour (basic case) is more expensive compared to the adjusted prices of the reserved instances but less expensive than the adjusted prices of the on-demand instances.

By comparing the cost per core hour for 2011 streaming from e-FISCAL calculations **(Basic Case)** and the Amazon **performance adjusted prices** for Cluster Compute Quadruple Extra Large instances - Linux per core hour we get the following results. The performance degradation used to adjust the prices of Standard M/L/XL instances is 43%



	e-fiscal median cost per core hour	e-fiscal average cost per core hour	Adjusted Price per core hour (€/hour) - 80% - 1year reserved	Adjusted Price per core hour (€/hour) - 60% - 1year reserved	Adjusted Price per core hour/ On demand
Cluster Compute Quadruple Extra Large - Linux	0.031	0.072	0.117	0.143	0.174

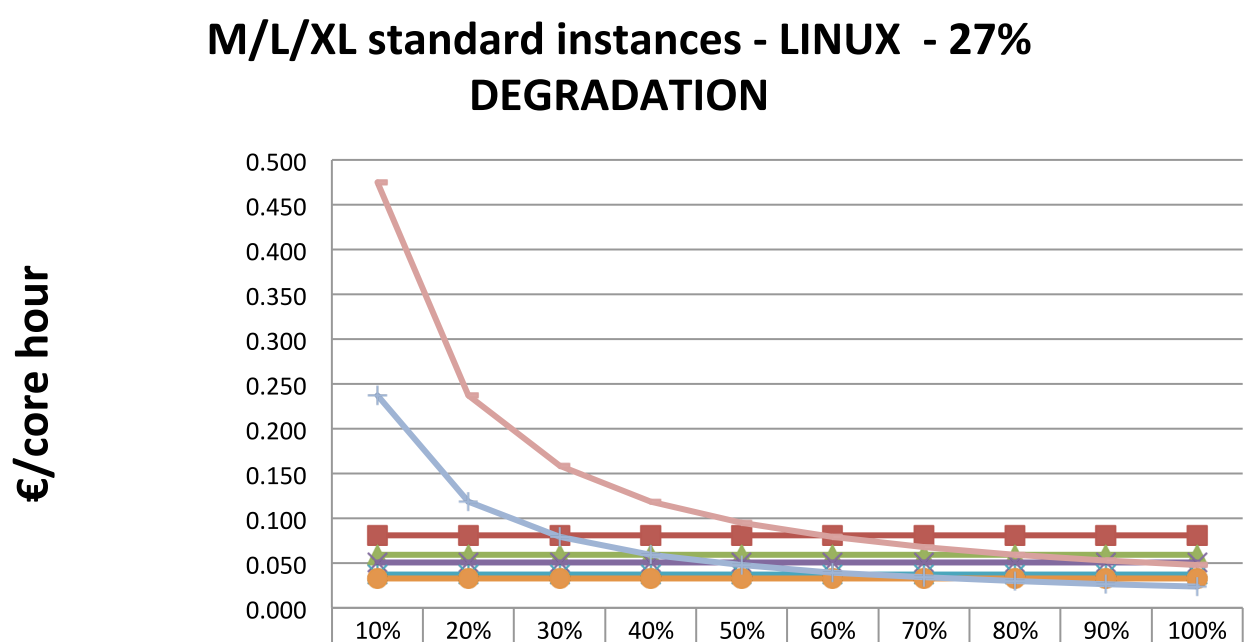
Graph 12: Comparing e-FISCAL cost with Amazon adjusted for performance prices (Cluster Compute Quadruple XL instances – Linux)

The median e-FISCAL cost and the average e-FISCAL cost (basic case) per core hour are less expensive than the adjusted prices of both on demand and one-year reserved instances.

4.4.3 Sensitivity analysis

Apart from the price comparison before and after performance adjustments we also conducted a sensitivity analysis.

Firstly we tried to approximate the indifference point between the cost of in-house computing provision and the performance adjusted Amazon prices by treating the utilisation rate of the infrastructure as a variable component. In the following graphs the cost per core (average and median) for 2011 for the total e-FISCAL sample is translated into cost per core hour under alternative utilisation rates; from 10%-100%. The cost per core hour is then compared with the prices of Amazon EC2 for on demand and reserved instances. In this exercise we assume that the cost per core per year remains the same regardless the utilisation rate. This is not absolutely correct as for example the electricity costs are influenced by the utilisation rate. Other costs could however are not expected to be affected by utilisation rate (e.g. premises costs, depreciation of e-Infrastructure, even the number of FTEs). We consider this process to be precise enough to come up with some rules of thumb that could permit centers to assess their efficiency vis-a-vis commercial cloud providers.



	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
On demand instances	0.081	0.081	0.081	0.081	0.081	0.081	0.081	0.081	0.081	0.081
1year reserved -60%	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059
1year reserved -80%	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050
3 year reserved -60%	0.037	0.037	0.037	0.037	0.037	0.037	0.037	0.037	0.037	0.037
3 year reserved -80%	0.033	0.033	0.033	0.033	0.033	0.033	0.033	0.033	0.033	0.033
e-FISCAL all median	0.237	0.119	0.079	0.059	0.047	0.040	0.034	0.030	0.026	0.024
e-FISCAL all average	0.475	0.237	0.158	0.119	0.095	0.079	0.068	0.059	0.053	0.047

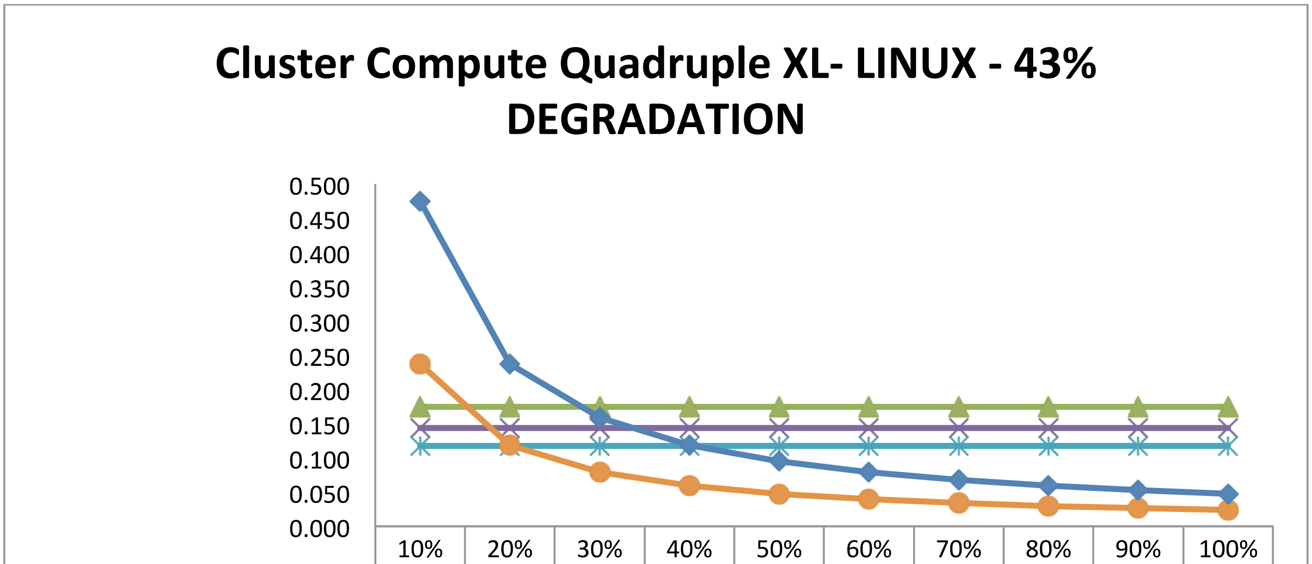
Graph 13: Identification of the Indifference point between e-FISCAL cost and Amazon adjusted for performance prices (Standard M/L/XL instances – Linux)

From the analysis in Graph 13, the cost per core/hour compared to the Amazon adjusted prices for Standard M/L/XL instances – Linux is as follows:
The cost per core hour of e-FISCAL median becomes less expensive when the e-Infrastructure operates:

- At 30% utilisation rate or more compared to on demand instances.
- At 40% utilisation rate or more compared to 1-year reserved instances utilised at 60%.
- At approx. 50% utilisation rate or more compared to 1-year reserved instances utilised at 80%.
- At approx. 65% utilisation rate or more compared to 3-year reserved instances utilised at 60%.
- At approx. 75% utilisation rate or more compared to 3-year reserved instances utilised at 80%.

The cost per core hour of e-FISCAL average becomes less expensive when the e-Infrastructure operates:

- At approx. 60% rate or more compared to on demand instances.
- At 80% utilisation rate or more compared to 1-year reserved instances utilized at 60%.
- At approx. 95% utilisation rate or more compared to 1-year reserved instances utilized at 80%.
- The cost per core hour of e-FISCAL average is more than the performance adjusted price for 3-year reserved instances in 60% and 80% usage rates.



	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
On demand instances	0.174	0.174	0.174	0.174	0.174	0.174	0.174	0.174	0.174	0.174
1year reserved -60%	0.143	0.143	0.143	0.143	0.143	0.143	0.143	0.143	0.143	0.143
1year reserved -80%	0.117	0.117	0.117	0.117	0.117	0.117	0.117	0.117	0.117	0.117
e-FISCAL all median	0.237	0.119	0.079	0.059	0.047	0.040	0.034	0.030	0.026	0.024
e-FISCAL all average	0.475	0.237	0.158	0.119	0.095	0.079	0.068	0.059	0.053	0.047

Graph 14: Identification of the Indifference point between e-FISCAL cost and Amazon adjusted for performance prices (Cluster Compute Quadruple XL instances – Linux)

From the analysis in Graph 14, the cost per core/hour compared to the Amazon adjusted prices for Cluster Compute Quadruple XL instances – Linux is as follows:
The cost per core hour of e-FISCAL median becomes less expensive when the e-Infrastructure operates:

- At approx. 15% utilisation rate or more compared to on demand instances.
- At approx. 20% utilisation rate or more compared to 1-year reserved instances utilised at 60% or 80%.

The cost per core hour of e-FISCAL average becomes less expensive when the e-infrastructure operates:

- At approx. 25% utilisation rate or more compared to on demand instances.
- At approx.30% utilisation rate or more compared to 1-year reserved instances utilised at 60%.
- At approx. 40% utilisation rate or more compared to 1-year reserved instances utilised at 80%.

It should be noted that the average and median utilisation rate for 2011 of e-FISCAL sample was 65% and 75% respectively.

4.5 Research question 4: Business models

This section is devoted to investigate possible scenarios for business models that envision a combined usage of public-funded federated infrastructures with commercial cloud providers. Furthermore, it will derive pricing models characteristics being used for cloud services so as to organize concepts and support e-Infrastructure managers in defining possible pricing schemes for their services.

4.5.1 Integration of e-Infrastructures with commercial cloud providers

As mentioned, commercial cloud services are becoming appealing for the research sector, especially in low-end computing. Additional added benefits include access to different types of resources/features or service levels and more elasticity. The role of Cloud Service Brokers (CSB) is emerging in different sectors, which is also being highlighted in Europe through the Helix Nebula project (http://www.helix-nebula.eu/) and the EGI Federated Cloud (http://www.egi.eu/infrastructure/cloud/). These initiatives are answering questions and putting into practice a number of cloud services dedicated for researchers while showing how federated infrastructures can work alongside and/or integrate with commercial providers. This can be shown through a number of scenarios that we will develop in this section. All scenarios consider that academic users relying on a publicly funded federated infrastructure with needs to expand the resource usage to commercial providers. The first scenario (Figure 10) envisions a peering between the public funded infrastructure broker with an external broker service that aggregates commercial cloud providers (e.g., as being designed by the Helix Nebula Initiative).

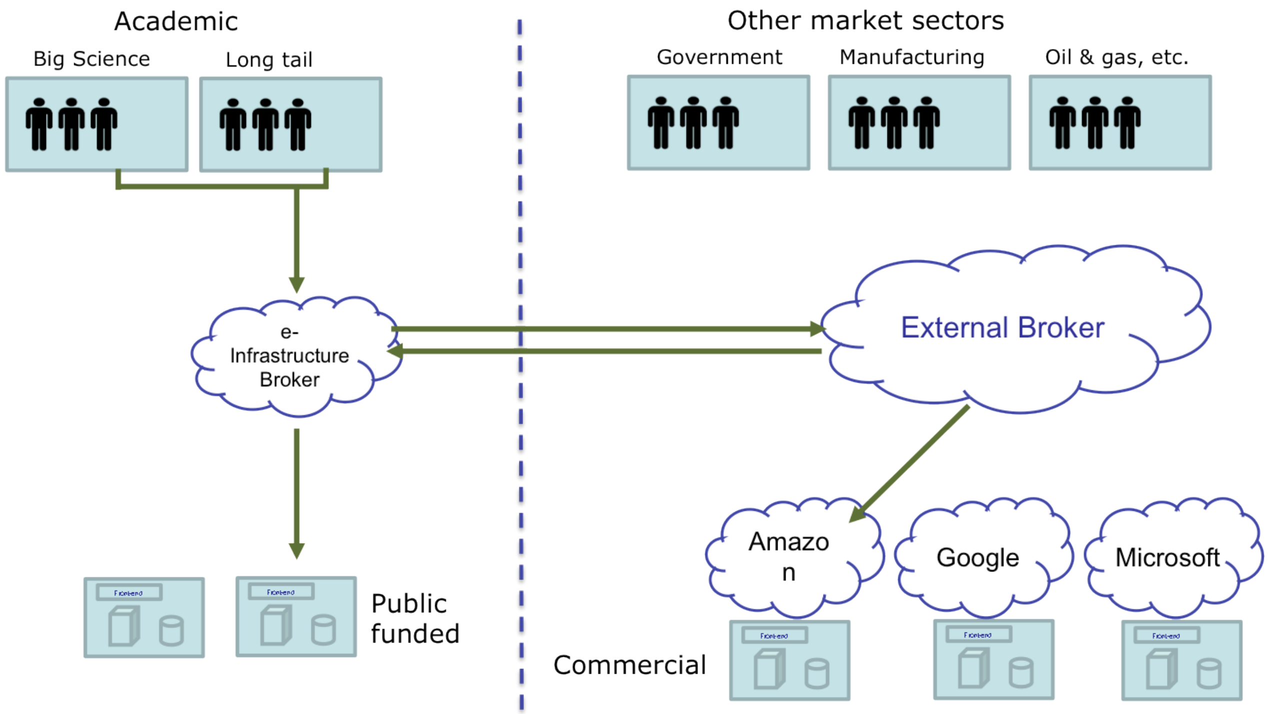


Figure 10: Federated Infrastructure Bursting

The second scenario is a direct usage of an external broker from an academic user (Figure 11). This may happen if the external broker offers extra capability that the public funded infrastructure broker is not able to offer. Single service providers in the public funded sector could integrate directly with the commercial broker so that an academic user could reach both public-funded and commercial cloud providers.

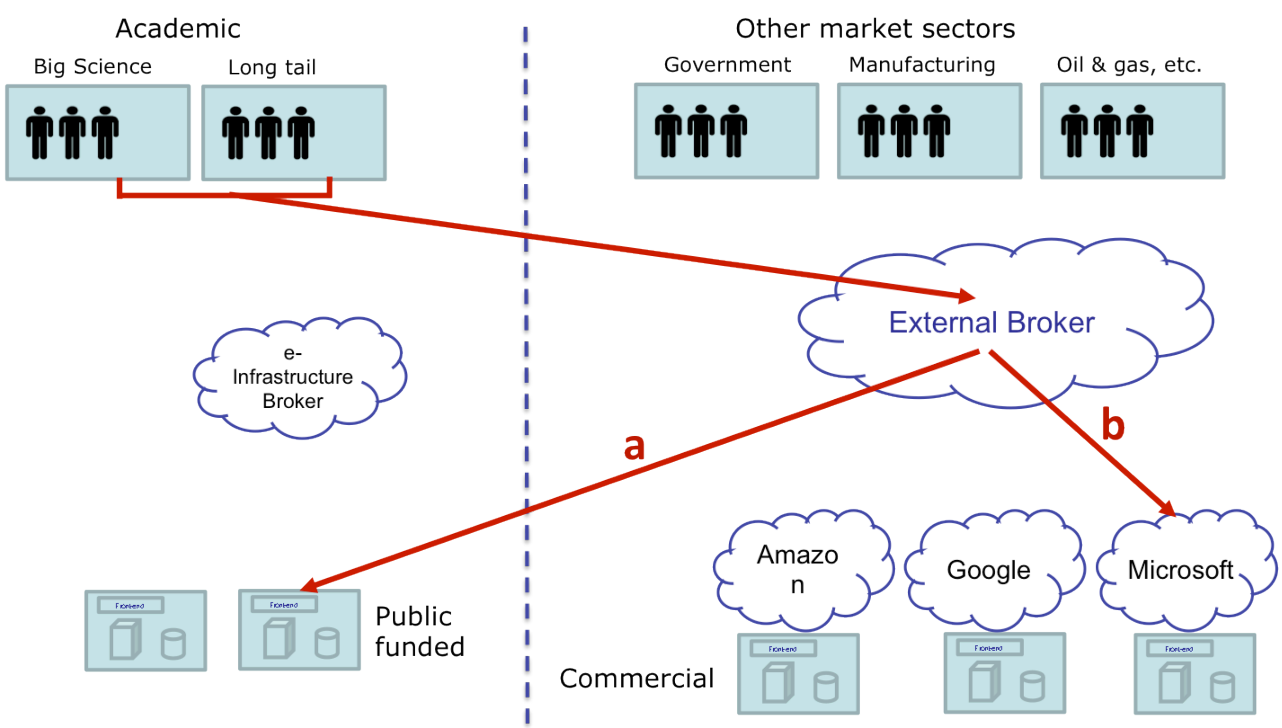


Figure 11: Integration with External Broker

The third and last scenario is the single provider bursting (Figure 12) that is the scenario where a public funded organisation providing ICT services to researchers aims to offer extra capacity that is not able to timely procure in house, thus integrating with an external provider to transparently move users' workload to there.

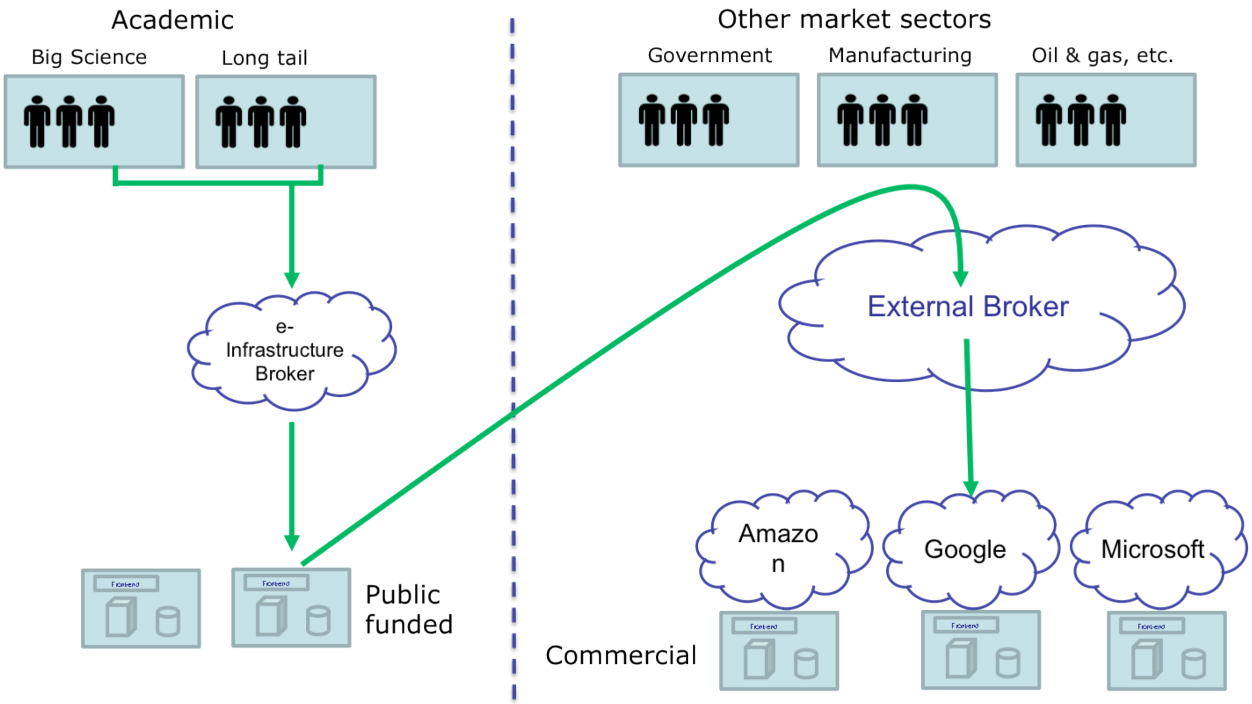


Figure 12: Single Provider Bursting

These integration scenarios are only possible if users own budgets to buy cloud services or if they are offered credits to use them. Usually, they access ICT services that are free at point of delivery, therefore the provision of commercial services need to be properly analysed also in terms of funding streams. Some questions remain open for the policy makers: how to reconcile the budget planning cycle for research computing to the pay-as-you go or subscription pricing models of cloud? What happens to the stored data if a budget is not renewed or is approved late? Who should own the budget? (e.g., research group, institution). There are many aspects that need to be considered also in the area of contracts and SLA's to make sure that the integration with commercial providers will create value for the academic users.

The last part of e-FISCAL questionnaire contained questions regarding cloud computing, Green IT aspects and sustainability considerations. The time horizon of these questions tried to balance recent past actions (i.e. actions that have been realized in 2011) and future prospects.

As for cloud computing, the most frequently used cloud computing service is Software as a Service (SaaS), followed by Infrastructure as a Service (IaaS). However the prospects for the future are in favour of a more intensive, compared to 2011, cloud computing services usage. Nevertheless, most respondents are more prone to use (free of charge) cloud computing services compared to those expressing their intention to buy such services. Finally the idea of imposing usage fees is not out of the question for several respondents that have included such provision in both the short and long term planning. A detailed analysis of the responses is found in Appendix 7.4.

4.5.2 Pricing Models

Many classifications have been proposed in the area of pricing schemes as presented in the state of art. The identified pricing models and classification that follow have been derived by analysing the real-world offerings from the major commercial cloud providers. Usually, all providers offer at least usage-based and subscription-based models, while currently only one provider digs into the market-based model. Strategy-based models may appear for certain customer domains (e.g. education, not for profit). Therefore, by analysing the current market for IaaS cloud services, we can derive the following schemes: Usage Based Pricing, Subscription, Market Based Pricing and Strategy Based.

The following table examines three of the main players in the IT sector: Amazon (A), Google (G), and Microsoft (M) and their adoption of the identified pricing models and their variables.

Variables	Description	A	G	M
Resources	Price depends on type of configuration (e.g., number of core, performance, RAM size)	✔	✔	✔
Features	Price depends on features (e.g., SLA, OS type)	✔	✔	✔
Region	Price depends on data center/geographical location	✔	✔	✖
Tier-based	Depends on segments of consumed time units	✔	✔	✔

Figure 13: Usage-based Model

Usage-based models (Figure 13) change the risk-sharing between the service provider and consumer by limiting the commitment from users as they can quit the service consumption any time without incurring in extra-costs. While this brings agility to the user, it can negatively impact on the cash flow of the service provider. That is why usually these pricing models are more costly than other options that imply a longer commitment from users.

Variables	Description	A	G	M
Resources	Price depends on type of configuration (e.g., number of core, performance, RAM size)	✔	✔	✔
Features	Price depends on features (e.g., SLA, OS type)	✔	✔	✔
Region	Price depends on data center/geographical location	✔	✔	✖
Usage volume	Price depends on volume; higher volume commitment leads to lower price	✔	✔	✔
Overage	Price changes if exceeding usage	✔	✔	✔

Figure 14: Subscription-based Model

Subscription-based models (Figure 14) require a contract from users and a commitment to buy the service for a defined time period that may vary from months to years. This is more welcomed from suppliers who can better develop capacity planning.

Variables	Description	A	G	M
Auction	Buyers bid in increasing increments of price	✔	✖	✖
Market Based	Large number of buyers and sellers indicate their preferred price, but cannot influence it individually	✔	✖	✖

Figure 15: Market-based Model

Market-based models (Figure 15) are in their early phase in the cloud market. Among the leading providers, only Amazon offers this pricing model enabling users to resell capacity bought via a subscription-based model with an auction style approach.

Variables	Description	A	G	M
Penetration pricing	To target market segments very sensitive to price			
Skim pricing	To target market segments relatively insensitive to price	✖	✖	✖

Figure 16: Strategy-based Model

Strategy-based models (Figure 16) are usually adopted to gain market share and reach customers that are, e.g. very sensitive to price or very insensitive. Leading providers can use penetration price to enlarge their user base in the science area (e.g., Google Apps offered for free to education).

From the analysis, we can observe that IaaS has complex pricing model if compared to SaaS, with PaaS is still developing. Also, computing services in the IaaS layer have more complex pricing model than storage/network with richer configuration options. With the evolution of the market, pricing models for compute services are expected to become simpler and with differentiation happening at the level of support, SLA or performance.

4.6 Research question 5: Green IT initiatives in HPC and HTC centres

The research question 5 was not one the main focus of our study as other studies are solely dedicated to such purposes. As an example GÉANT Green Team64, ECONET65 and others on Green IT were presented in the e-FISCAL workshops as case studies. However, we made an effort to map the main activities in which the HPC and HTC centres are involved regarding Green IT initiatives. We also gathered information about PUE numbers that has become one of the most significant metrics in relation to Green IT (Cordis, 2011).

[64] http://www.geant.net/Network/Environmental_Impact/Pages/home.aspx
[65] http://www.econet-project.eu/

On the basis of the responses gathered, respondents proved very positive towards Green IT initiatives and some of them had already embarked into projects towards this direction66. Energy efficiency considerations had played a role in investment decision of almost all respondents (24 out of 26 answers) while changes aiming at increasing energy efficiency were applied in more than half the respondents (14 out of 23). Finally, ten out of twenty-four respondents reported that in 2011 had recycled CPUs and storage devices.

Please answer the following questions in relation to "Green IT"		
Answer Options	**Yes**	**No**
Did you recycle CPUs or storage devices during 2011?	10	14
Do you plan to use some sort of "Green IT" in the future?	24	2
Had energy efficiency considerations influenced your acquisition decisions in 2011?	23	2
Did you make any changes in your hardware/software environment to increase the energy efficiency in 2011?	14	9
Did you use any form of "Green IT" in 2011?	14	8

Table 13 – Green IT related actions in 2011 and future prospects

Respondents generously provided details on their plans and experiences with Green IT. The majority proved very positive towards Green IT initiatives and some of them had already embarked into projects towards this direction. Examples of Green IT initiatives reported by respondents are summarized as follows:

- Buying energy efficient servers (improve performance per Watt).
- Virtualising more IT services.
- Reusing heat from servers to warm water for nearby buildings.
- Buying new hardware to replace old hardware.
- Building new datacentres.
- Appling efficient cooling systems.
- Exploitation of external temperature in order to use free cooling, fully or partially, during the whole year.
- Machine rooms in the national infrastructure capture/recycle heat from the compute systems.
- Reallocation of HPC systems. Such decisions are partially driven by Green IT considerations (no use of fossil energy sources) as well as by cost considerations (low electricity cost).
- Improvement on airflow management by eliminating mixing and recirculation of hot equipment exhaust air and maximizing the return of air temperature to the computer room air-conditioning unit.
- Implementation of environment monitoring systems in order to better measure resources.

Adding additional points in tendering processes if the use of energy is better related with performance.

As far as PUE numbers are concerned the average and median PUE for 2011 of the sample sites were 1.56 and 1.50 within a range of 1.25 (minimum value) to 2.24 (maximum value). Apart from PUE ratio improvement from 2010 to 2011 these numbers are aligned with best practice.

[66] It should be noted that as reported in Cordis report (2011) the European Commission is highly interested in new approaches to minimising cooling requirements, to reducing power consumption and increasing the re-use of heat generated.

5. Discussion and Recommendations

5.1 State of the art

The state of the art review brought significant inputs and insights into the e-FISCAL project, from costing models, categories and methodologies to business and pricing models. Most importantly it provided clear case studies in the US and UK whose numbers could be easily be compared with the e-FISCAL ones, as long as the fact that these case studies were ranging from single sites to a national infrastructure was kept in mind. The state of the art became a focal point of e-FISCAL and gathered considerable part of the web site traffic.

5.2 Methodology and cost model

As a follow-up from the first point (state of the art review), it is evident that e-FISCAL has been a pioneer effort trying to capture overall picture of the highly distributed and heterogeneous European environment. The e-FISCAL hybrid methodology and cost model were realistic so as to be able to provide the expected results. The pros and cons of the e-FISCAL model are summarised below.

The merits:

- It is a **hybrid** model that builds on FCA and TCO and adapts to real case constrains.
- It includes all **cost categories** identified in literature are relevant to e-Infrastructures costing.
- It balances easiness of information collection with precision in results.
- It is easy to apply as it only has a few input requirements.
- It provides acceptable cost estimations precision. It does not aim at providing detailed costing data but is good enough to approximate cost. It is also suitable for cross site comparisons and cost assessments through time.
- It is transparent and auditable.
- It is suitable to perform sensitivity analysis.
- It can be used by stakeholders outside the organization.

However the limitations of the model are as follows:

- It cannot generate as precise cost estimations for a specific centre as Full Cost Accounting (provided that all costs are accounted for in accounting books).
- The simulation of e-Infrastructures is mainly based on current or recent procurement costs. As prices go down (our evidence confirms that, see Del. 2.2), we may have underestimated CAPEX (if infrastructure in place has been procured in higher prices). Nevertheless, old machines consume more energy which balances things. OPEX has been correctly estimated regarding electricity consumption.
- There is no differentiation in the calculated cost per core between in-house machines. We treat core hour as a homogenous service regardless of the machine.
- The accuracy of findings is reliant on the completeness and reliability of inputs. Nevertheless, the interviews provided clarifications where issues have been identified and we have also isolate outliers.
- The model concentrates on computing cost not on storage. However, part of the storage cost is included (as with other similar studies).

5.3 e-FISCAL sample and main findings

e-FISCAL received 28 high quality answers from 16 countries which constitute a diverge sample from different types of centres in terms of size (small and large), in terms of nature (HPC and HTC), with a low or high number of staff, etc. e-FISCAL was able to calculate several metrics, including cost per core hour, CAPEX/OPEX ratio, depreciation rates, cost distribution, electricity PUE. The main findings can be summarised as follows:

- **Cost effectiveness**: In-house HPC/HTC e-Infrastructures are cost-effective with the high average utilisation rates (65-75%) and depreciation rates (5 years) recorded. However, per application cost analysis is needed and e-FISCAL did not have the resources to execute several comparative application case studies.
- **Cost structure**: OPEX dominates the costs (around 70%) – Personnel costs are around half of the total costs, while personnel costs vary a lot around Europe.
- **Comparison with state of the art:** The e-FISCAL findings are in-line with the results of similar studies identified in the state of the art through literature review.
- **HPC- HTC comparison:** It appears that the HPC sites of the e-FISCAL sample (national Tier-1 centres) are in average cheaper than corresponding HTC ones. Possible reasons for this are the higher number of cores per CPU, the state of the art energy systems, and the fact that larger sites have in most cases less FTEs per core.
- **Cost on the basis of size (no of cores) -** Big vs. small sites**:** It appears that sites with <1000 cores are much more expensive than sites >1000 cores. Indicative values for cost per core hour are around € 0.09 for the first category and € 0.05 for the latter.
- **Cost structure heterogeneity:** there is heterogeneity in the cost structure of sites/centres (in terms of cost categories). However, it is rather evident that in the vast majority of cases OPEX is not less than 40% of total costs and in the average case OPEX dominates costs.
- **In-house HTC and HPC vs. commercial Clouds; Benchmarking:** Cloud service providers usually have a wide portfolio of instances and services.
    - Comparisons between in-house and commercial clouds have to be made very carefully, trying to compare like with like. Benchmarking is thus very relevant in this exercise as it identifies comparable performances between in-house configurations and cloud instances, especially because the cloud instance description might not always be easily comparable.
    - Performance degradation of cloud virtual machines with the in-house e-Infrastructure have to be accounted for through appropriate weight factors, especially in cases where virtual machines are compared to bare metal.
    - With the small-scale benchmarking exercise a 27% and 43% degradation has been averaged for HTC and HPC respectively.
    - Taking these degradations into account and comparing Amazon EC2 with the e-FISCAL HTC and HPC costs, it is obvious that:
        - Amazon standard reserved instance prices are comparable with e-FISCAL ones and sometimes cheaper (the 3-year reserved ones)
        - Amazon standard On-demand ones are always more expensive than e-FISCAL
        - Amazon cluster compute ones (HPC in the cloud) are always much more expensive (1.5 to around 5 times)
            - Several disclaimers have to be taken into account though as analysed in the previous section such as comparing 2012-2013 prices with 2011 costs, the exclusion of

networking and storage costs, the varying utilisation rates and the fact that it might not be possible to absorb all the cloud reserved capacity and more.

- Utilisation is a crucial parameter and thus a sensitivity analysis has been performed showing the indifference points between in-house and clouds.
- However a general conclusion is that the Cloud is not always cheaper as it might have been perceived a few years back, especially in shared and well utilised e-Infrastructures.

- **Business and pricing models:**
    - The e-FISCAL work can help e-Infrastructures understand the cost element, while a state of art analysis of pricing schemes for cloud can further support them in identifying possible pricing approaches for services if service fees are applied to users in the future.
    - In the framework of the e-FISCAL project, we have elaborated business models for federated infrastructures and understood how commercial cloud providers could be integrated in mixed usage scenarios.
    - We have also analysed the pricing models characteristics being used for cloud services so to organize concepts and support e-Infrastructure managers in defining possible pricing schemes for their services.
- **Green IT:** Given the fact that electricity is an important cost factor (around 15% of the total cost), Green IT practices are continuously being adopted in order to minimise the PUE as close to one. Several "clever" practices are being deployed to minimise costs and rely on Green energy as described above and this trend will continue to grow in the future.

5.4 e- FISCAL main contributions

e-FISCAL main contributions can be summarised as follows:

- e-FISCAL contributed to the **appreciation of cost aspects** by the community and the fact that nothing is free
- e-FISCAL was **pioneer to estimate cost ranges for European e-Infrastructures** at this scale and contributed in a good enough understanding of e-Infrastructure cost ranges
- Similar contribution was made with regards to the understanding **how in-house publicly funded e-Infrastructures are compared to commercial cloud** ones and benchmarking played a crucial role in arranging as much as possible like with like comparisons
- e-FISCAL has contributed to the community **free self-assessment tools**, both as a downloadable file and as a web-based application that can be used to assess the cost of new or existing sites (www.efiscal.eu/tools)
- e-FISCAL has **created a community of experts** and interested in cost aspects with around 50 financial experts and 150 interested ones, organising 4 well attended workshops.

5.5 e- FISCAL recommendations

Further general observations, including strengths or weaknesses that may require recommended actions are the following:

- Cost per core hour is not always accurate, however it is probably good enough to get a good understanding of computing costs.
  - Alternative approaches should be also attempted such as per application or use case cost analysis.
- Comparing in-house costs with commercial cloud prices is a first good step in understanding how the two relate. Evidence so far indicates moderate savings in personnel costs when moving to the cloud (based on sporadically answered questions in the e-FISCAL survey).
  - Migration to the cloud deserves a different analysis focusing on avoidable costs and time value of money.
- Decisions are not only based on costs. Strategy and policy considerations play a significant role. Other important non-financial aspects include speed to adjust to increases in capacity demands, the easiness and willingness of porting applications from one environment to another, security considerations as well as the facilitation of knowledge transfer and expertise related to leading edge ICT challenges.
  - Costs should be well understood, however there may be other priorities (strategies or policies)
  - Cost is different from price; profit margin or even loss (loss leader strategy) need to be kept in mind. And cost is different from value.
  - In all cases, besides quantitative, qualitative aspects should be examined (e.g. quality of service and value).
- Studying business and pricing models from commercial cloud providers and understanding costs also helps identify suitable business and pricing policies for e-Infrastructure providers.
- On business models:
  - high utilisation is key to cost efficiency: cloud service brokers should be adopted as a way to aggregate demand and maximise utilisation of federated infrastructures
  - integration between e-Infrastructures and commercial providers should be pursued to expand available capacity to the user communities, variety of resources, service differentiation, freedom of choice
  - value added services should be identified (e.g., backup, compliance, specialised user support) and marketed to better communicate the value differentiation on top of the basic features
- On pricing models:
  - for organisations aiming to directly charge users or their institutions for computing services, at least two types of pricing models should be defined:
    - usage-based pricing scheme to meet the need of elastic demand
    - subscription-based pricing scheme to meet the need for longer term commitment with lower price; different durations should be explored with the demand side

The main recommendations are summarised **per stakeholder**:

- **e-Infrastructure computing providers and HPC/HTC centre managers**

- e-Infrastructure computing providers and HPC/HTC centre managers should keep track of their cost structures and their costs through a systematic way to assist in planning and determining their centres' sustainability.
- e-Infrastructure computing providers should be making their costs available openly contributing to the Open Data and Open Science paradigm, especially as the related funding is based on public funding.
  - The e-FISCAL tools are open freely for those that don't have related methodology and model

- **Research communities and end users**
  - As computing costs is not the sole issue in the in-house vs. commercial cloud discussions, users should make an effort to collect and document the experiences on the use of in-house vs. commercial clouds in order to get a more holistic view of the effort and complexity required in doing their jobs
- **e-Infrastructure policy makers and funding agencies**
  - Funding agencies of research infrastructures should request from all e-Infrastructure providers that receive public national or European funding to ensure publication and open access to data about funding of their e-Infrastructures[67].
  - Given the uprising competition from commercial cloud providers, e-Infrastructure policy makers and funding agencies should recognise the cost parameter as an important parameter in the sustainability of computing e-Infrastructures.

5.6 e- FISCAL sustainability and exploitation

The sustainability of the e-FISCAL approach depends to a high degree on the number of individuals and organisations that adopt it in the future activities. The survey results themselves remain relevant only relatively short period, but helping European e-Infrastructure actors to align their cost assessment and reporting methodologies in terms of cost categories and even tools used to estimate and gather information will have a considerably longer positive impact. To achieve these goals, it is essential that they are both continuously used in a sufficiently visible project and made available under licensing terms that maximise the uptake.

The first prerequisite is fulfilled through the adoption of the e-FISCAL survey methodology by the annual **EGI compendium efforts**. The survey will collect wide range of information related to the EGI infrastructure and makes this data available in the pan-European compendium document. The next survey will include the full cost-related questionnaire developed in the context of the e-FISCAL project.

The **e-FISCAL tools** themselves will remain available through the e-FISCAL project website. The cost models are published in the project deliverables, and the tools used to analyse the information are licensed in a way that maximises the uptake while still ensuring that the link to the e-FISCAL project remains visible. To achieve this goal, the following licencing approaches have been chosen:

[67] Based on the e-IRG related recommendation from the e-IRG paper: Cloud Computing for research and science: a holistic overview, policy, and recommendations http://www.e-irg.eu/images/stories/dissemination/e-irg_cloud_computing_paper_v.final.pdf

- The online tool is released under Affero General Public License, version 1. This allows anyone copy the online tool, modify it and install and provide it (either in original or modified form) as a service on their own environment **providing** that the source code used to provide service will be made available on the same server that is running the software.
- Other tools, such as the Excel-based cost assessment tool, are released under the "Creative Commons Attribution ShareAlike 3.0 Unported" license[68]. This license grants anyone the right to share (copy, distribute) and adapt (including for commercial purposes) the work, **provided** they mention that the work is based on e-FISCAL contributions and that any derived work is distributed under similar license. We also have developed a questionnaire with ranges that we did not have time to exploit. This can be also considered as one extra tool.
- The e-FISCAL consortium partners are willing to provide consulting services to computing centres or even other e-Infrastructure components, as well as to other research infrastructure (non-e-Infrastructures such as ESFRI ones). An alternative method can be to partner in proposals or projects with such related entities.
- The e-FISCAL consortium expertise and know-how can be further exploited to understand better the cost of data infrastructures and data management, although being more complex and other methodologies should be more appropriate.

5.7 Discussion: Applying e- FISCAL results in managerial decision making

In the following sections the project findings are put into the context of strategic or operational managerial decision making in an organisation using non-trivial[69] amounts of IT resources to support its activities. The cost-related mechanisms influencing sustainability of non-profit organisations may be more indirect and less immediate in their nature than ones in the commercial marketplace, but in both cases cost-effective, reliable and efficient provisioning of services that fulfil the current and future needs of the end-users is a crucial long-term success factor for any organisation.

5.7.1 Cost-effectiveness considerations

The analysis of the cost effectiveness based on the e-FISCAL results indicate that an in-house infrastructure having high utilisation rate (e.g. over 65%) is in almost all cases cost effective when compared to cloud pricing. This assumes that the personnel costs in the overall yearly IT expenditure are roughly 50%, i.e. similar to average/median values in the e-FISCAL study.

In the case where the organisation uses (mainly) in-house IT services and the utilisation rate is high, the cost saving potential is rather limited. Rather than focusing on IT cost savings the attention should be focused on variations in the demand and estimating the probability and business impact of situations where the demand for IT resources exceed the in-house capacity. If either one of them is high, investigating whether additional in-house capacity or development and deployment costs of a hybrid infrastructure (including Cloud resources to handle

[68] http://creativecommons.org/licenses/by-sa/3.0/

[69] It is difficult to quantify universally what the "non-trivial" means since it depends quite heavily on the nature of the organisation.

peak loads) could be justified by the additional revenues or other benefits (such as increased user satisfaction leading to improved retention of clients).

On the other end of the scale, in case the utilisation rate of the in-house infrastructure is very low (e.g. 10%), it is most likely a good idea to consider outsourced solutions. However, there are two issues to consider in the process:

- What the actual savings potential is and when it can be realised. Avoiding the purchase of new servers brings in savings, however some investments (e.g. related to building a server room) tend to be sunk costs, i.e. impossible to recover despite early decommissioning. In case the sunk costs dominate, studying the ways the productive utilisation rate of the in-house infrastructure could be increased might be useful alternative strategy. For example contributing resources for educational or charitable purposes is a relatively common marketing strategy that could be applied.
- Whether the outsourced solution fulfils all the functional and non-functional requirements, and whether lack of in-house infrastructure will have a negative impact on staff motivation and skills, or on development, training and support activities. For technical performance, consider benchmarking.

Considerations for an organisation using (mainly) Cloud resources the situation is perhaps slightly more straightforward. Investing in in-house IT infrastructure makes sense in situations where there is sufficient volume and it is possible to identify a baseline load that seems to remain stable. The potential benefits are most obvious in cases where the volume of activity is very high and the business model of the organisation makes cost of service provision critical in terms of competitiveness or profitability.

The case of in-house IT infrastructures that fall into neither of the extremes require case by case analysis that usually needs a broader analysis than just review of costs of different options.

5.7.2 Cost structure, its heterogeneity and size of the infrastructure

The e-FISCAL study indicated that OPEX costs dominate, with the personnel costs typically being the largest category. In the e-FISCAL material this was true even in large scale infrastructures (although other studies have shown cases of large-scale installations where the personnel costs have been a relatively minor cost category).

Based on the results, scaling up may sometimes brings in economies of scale, however, case-by-case analysis is important. Once moving beyond very limited IaaS offering with (explicitly) minimal support contract, the actual applications and services provided may change the cost per core hour considerably (i.e. difference between offering IaaS solution and IaaS offering with support and consulting). Especially when launching new services or entering new markets, it may be difficult to estimate how much of additional manpower is needed in marketing, user support and consulting to gain critical mass. Based on literature review switching to an outsourced solution does not seem to bring in major savings in manpower expenditure, even on relatively "technical" management tasks.

The ambiguous results of the "economies of scale" analysis as well as lack of demonstrated dramatic manpower savings may be indications of the relatively low maturity level of the IaaS marketplace. Standards and best practices are evolving quite rapidly and "standard categories of use cases" are still in the process of forming, leading to uncertainty and management overhead on the "client side". Development in these areas should be monitored, as they may change the cost structure of outsourced solutions while at the same time making the overall costs more predictable.

5.7.3 Green IT

Any organisation providing services where IT is in a visible role it is becoming advisable to show awareness of the basic metrics related to Green IT. Being able to discuss PUE value of the infrastructure used (whether in-house or outsources) is most likely becoming a relevant corporate social responsibility and marketing issue. Monitoring the developments of more multi-faceted Green IT metrics and initiatives is also an important trans-disciplinary area where the communications, business and IT experts should all be involved.

When considering establishing large-scale in-house infrastructures, developments related to compliance and regulatory issues may become an issue that should be taken into account as part of the cost comparison between different options as well as capacity planning.

5.7.4 Specific considerations for SMEs

In the smaller SMEs some of above steps might be difficult to implement. It may be that (at least in the initial stages of the company) there are no dedicated IT staff members, or the ways to track the use of their time may not be suitable for analysis of the IT-specific issues (e.g. if the reporting is based on client projects instead of IT management functions).

However, even in this case the e-FISCAL results can be used to estimate costs for planning purposes. The tools and supporting material can also be used to show awareness of IT cost issues with minimal effort and in a way that is grounded on published methodology. For example when negotiating with new investors this would ideally allow both parties to focus on business strategy and revenue models.

6. REFERENCES & RESOURCES

Carlyle A. G. et al, (2010), "Cost-effective HPC: The community or the cloud", IEEE Conference on Cloud Computing Technology and Science.

Cohen, S. and Fotis Karagiannis, e-IRGSP2 deliverable D2.4 available at www.e-irg.eu/publications

Cordis (2011), Workshop Report "Towards A European Roadmap for Datacentres and their Economic and Environmental Impact", held in Brussels in November 2011.

Denne, M. (2007). Pricing utility computing services. International Journal of Web Services Research, Vol. 4, No. 2, pp. 114-127.

EGEE, (2008), "An EGEE comparative study: Grids and Clouds - Evolution or revolution", Members of the EGEE-II Collaboration, 30/05/2008.

EGI-InSPIRE (2012), EU deliverable: D2.30, "Seeking New Horizons: EGI'S role in 2020" Members of EGI-InSPIRE Collaboration, 27/4/2012.

EGI-InSPIRE EU deliverable D2.6 - Integration of clouds and virtualisation into the European production infrastructure, Members of the EGI-InSPIRE Collaboration, 1/3/2011 (2011a).
(https://documents.egi.eu/public/RetrieveFile?docid=258&version=8&filename=EGI-D2.6-258-final.pdf)

EGI-InSPIRE EU deliverable D2.7 - EGI sustainability plan, Members of the EGI-InSPIRE Collaboration, 31/03/2011 (2011b)

(https://documents.egi.eu/public/RetrieveFile?docid=313&version=11&filename=EGI-D2.7-final.pdf)

Hammond, M., R. Hawtin, L. Gillam and C. Oppenheim (2010), "Cloud Computing for research", Final Report, Curtis + Cartwright Consulting Ltd (7/7/2010).

Harmon, R., Demirkan, H., Hefley, B. & Auseklies, N. (2009). Pricing Strategies for Information Technology Services: A Value-Based Approach. Proceedings of 42nd Hawaii International Conference on System Sciences, pp. 1-10.

Hawtin et al. (2012), "Cost analysis of cloud computing for research", Final Report to EPSRC and JISC, 22/2/2012, Curtis+Cartwright Consulting Ltd.

HEP-SPEC Benchmark webpage, http://hepix.caspur.it/benchmarks/doku.php.

Hilton, R., M. Maher and F. Selto, "Cost management: Strategies for Business Decisions", International Edition, 4th edition, Mc Graw Hill, 2008

Hopkinson, L. and P. James, "ICT Related Energy Use, Costs and Carbon Emissions in UK Universities and Colleges - Results from Use of the JISC/SusteIT Footprinting Tool, April 2012, University of Bradford and SusteIT.

Kashef and Altmann, (2012), "A cost model for Hybrid Clouds"in K. Vanmechelen, J. Altmann, and O.F. Rana (Eds.): GECON 2011, LNCS 7150, pp. 46–60, 2012.

Koomey et al. (2009), "Assessing trends in the electrical efficiency of computation over time", CA: Analytics Press, Oakland.

Koomey et al. (2011), "Growth in Data centre electricity use 2005 to 2010", CA: Analytics Press, August 1st 2011, Oakland.

Lin A. and Chen (2012), "Cloud computing as an innovation: Perception, attitude, and adoption" International Journal of Information Management, http://dx.doi.org/10.1016/j.ijinfomgt.2012.04.001.

Magellan Report on Cloud Computing for Science (2011), U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research (ASCR), December, 2011 (http://science.energy.gov/~/media/ascr/pdf/program-documents/docs/Magellan_Final_Report.pdf)

Marston et al. (2011), "Cloud computing — The business perspective" Decision Support Systems, Vol. 51, pp. 176–189.

Misra, S.C and A. Mondal, (2011), "Identification of a company's suitability for the adoption of cloud computing and modelling its corresponding Return on Investment", *Mathematical and Computer Modelling*, Volume 53, Issues 3-4, pp. 504-521.

NAS Parallel Benchmark Technical Report, available at http://www.nas.nasa.gov/assets/pdf/techreports/1994/rnr-94-007.pdf.

Nazir, A, S- A. Sørensen (2010), *"Cost benefit analysis of high performance computing infrastructures"*, Paper presented in Service Oriented Computing and Applications (SOCA) 2010 IEEE International Conference.

Niemi, T., & Hameri, A. (2012). : Memory-based scheduling of scientific computing clusters, J Supercomput (2012) 61:520–544

Niemi, T., Kommeri, J., & Hameri, A. (2012). Improving Energy-Efficiency of Scientific Computing Clusters. In N. Kaabouch, & W. Hu (Eds.), *Energy-Aware Systems and Networking for Sustainable Initiatives* (pp. 1-19). Hershey, PA: Information Science Reference.

Opitz, A., H. König and S. Szamlewska, (2008), "What Does Grid Computing Cost?" *Journal of Grid Computing*, Vol. 6, No. 4, pp. 385-397.

Osterwalder, A. "Business Model Generation", Wiley 2010

Palaiologk, A. A. Economides, H. Tjalsma and L. Sesink (2012), "An activity-based costing model for long-term preservation and dissemination of digital research data: the case of DANS" Int. J. Digit Libr, Vol. 12, pp. 195–214.

Paleologo, G. (2004). Price-at-Risk: A methodology for pricing utility computing services. IBM Systems Journal, Vol. 43, No. 1, pp. 20-31.

Rasmussen, N. (2011), "Determining Total Cost of Ownership for Data Center and Network Room Infrastructure, White paper 6, Schneider Electric.

Risch, M and J. Altman, "Cost Analysis of Current Grids and Its Implications for Future Grid Markets", J. Altmann, D. Neumann, and T. Fahringer (Eds.): GECON 2008, LNCS 5206, pp. 13–27, 2008.

Smith, P., (2011) A Cost-Benefit Analysis of a Campus Computing Grid, *Purdue University*

University of Washington "Hyak Operating Costs and Comparison with Commercial Alternatives", http://escience.washington.edu/get-help-now/hyak-operating-costs-and-comparison-commercial-alternatives (accessed 30/11/2011)

Walker, E. (2009), "The real cost of a CPU hour", *Computing Practices*, Volume 43, Issue 4, pp. 35-41.

Willard, C, A. Snell, L. Segervall, HPC Budget Allocation Map: Industry Averages, Intersect 360 Research, March 2012.

Willard, C, A. Snell, L. Segervall, HPC Budget Allocation Map: Industry Averages, Intersect 360 Research, January 2013.

7. APPENDICES

7.1 e-FISCAL questionnaire (pdf)

Below is the e-FISCAL questionnaire based on the SurveyMonkey tool. By clicking on the image the pdf document opens (requires acrobat reader).

e-FISCAL questionnaire

1. e-FISCAL survey overview

The e-FISCAL project will build a cost model using anonymised real data to create a total yearly cost of ownership measure. This model will allow service providers and user communities to identify areas where the overall cost efficiency of ICT-enabled research can be optimised and provide useful input for their long-term sustainability planning. It assists planning by identifying the e-Infrastructure cost factors on a cost item basis. The scope includes analysing the costs and cost structures of the European High-Throughput and High-Performance Computing (HTC and HPC) and comparing the costs and the cost structures of these research e-Infrastructures with similar commercial leased or on-demand offerings. The comparison would go beyond a simple "cost per core hour" comparison by identifying and analysing the qualitative differences in service (such as quality of service and availability) between HTC or HPC e-Infrastructures and their closest commercial counterparts. A more detailed description of the project as well as an overview of the model are found in appendix A.
The questionnaire is targeted not only to NGIs and national HPC coordinators, but also to individual HTC/HPC centres. This is due to the fact that some key information (such as energy or auxiliary costs) is usually available at such (HTC/HPC centre) level. The questionnaire has two main sections. The first covers the necessary input data for the calculation of the total yearly cost of ownership: amortized investment costs and operating expenses. Therefore there are questions referring to the investment in e-infrastructure elements (hardware, computing storage, interconnect equipment and auxiliary equipment) as well as operating expenses related questions (e.g. personnel, electricity, premises costs). The generic cost model used in the study accompanied by benchmarking metrics produced by the analysis of the data will become public on the project's website. The second section is related to the sustainability outlook and Green IT aspects where questions about the current and future use of services by commercial service providers are discussed.
Data provided will be compared with existing accounting data available at a central level (such as EGI.eu accounting portal). The answers given will be considered as strictly confidential and only statistically processed results that guarantee anonymity will be publishable. For this reason even partially completed questionnaires are valuable for the project. Furthermore, the data will not be made public even after the project lifetime. For any questions related to the survey please contact survey@efiscal.eu.
This effort is funded by the FP7 EC project e-FISCAL (www.e-fiscal.eu).

Your participation is highly appreciated!

2. General information

1. To which e-infrastructure is your institute's infrastructure part of ?

- ☐ NGI/EGI
- ☐ National HPC infrastructure/PRACE
- ☐ Both
- ☐ Other (please specify)

Page 1

7.2 e-FISCAL questionnaire (spreadsheet and on-line version)

The final e-FISCAL questionnaire can be found in several formats in the e-FISCAL web site under www.efiscal.eu/tools, namely:

- Downloadable spreadsheet
- Web-based tool

For the excel-based tool and other versions of the questionnaire (e.g. ranges-based), please contact the e-FISCAL team: http://www.efiscal.eu/contact

7.3 Detailed methodology steps to calculate costs

The detailed methodology used in order to come up with the cost per core and the cost per core hour is presented in the following table. In the same time the average and median values of the total e-FISCAL sample are shown in the last to columns in order to enhance the comprehensiveness of the way the model works.

Type of data	Data number	Data description	Average values e-FISCAL sample (amounts in €)	Median values e-FISCAL sample (amounts in €)
Input	(1)	Average Cost per core	275	218
Input	(2)	Average cost per Tape storage TB	88	78
Input	(3)	Average cost per Disk storage TB	493	270
Ratio	(4)	Ratio Tape TB/core	16%	0%
Ratio	(5)	Ratio Disk TB/core	25%	7%
Calculation	(6) =(4)*(2)	Tape investment / core	14,1	0,0
Calculation	(7) =(5)*(3)	Disk Investment / core	123,3	18,9
Calculation	(8) = (1)+(6)+(7)	Total investment (cores+Storage)	412,3	236,9
Input	(9)	% interconnection over investment	10%	9%
Input	(10)	% support contract costs over investment	6%	4%
Input	(11)	% auxiliary equipment over investment	18%	20%
Calculation	(12) =(9)*(8)	Investment value interconnection	41,23	21,32
Calculation	(13) =(10)*(8)	Investment value support contract costs	25,56	9,48
Calculation	(14) =(11)*(8)	Investment value auxiliary equipment	74,22	47,38
Calculation	(15) =(12)+(13)+(14)	Other investment	141,02	78,18
Calculation	(16) =(15)+(8)	Overall invested capital	553,35	315,08
Input	(17)	Depreciation rate for cores	4,9	5,0
Input	(17A)	Depreciation rate for tape storage	6,5	5,0
Input	(17B)	Depreciation rate for disk storage	5,6	5,0
Assumption	(17)	Depreciation rate for interconnection	4,9	5,0
Assumption	(17)	Depreciation rate for Support contract	4,9	5,0
Assumption	(17)	Depreciation rate for Auxiliary equipment	4,9	5,0

Type of data	Data number	Data description	Average values e-FISCAL sample (amounts in €)	Median values e-FISCAL sample (amounts in €)
Calculation	(18)=(1)/(17)	Depreciation cost logical cores	56,1	43,6
Calculation	(19)=(6)/(17A)	Depreciation cost Tape storage	2,2	0,0
Calculation	(20)=(7)/(17B)	Depreciation cost Disk storage	22,0	3,8
Calculation	(21)=(12)/(17)	Depreciation cost interconnection	8,4	4,3
Calculation	(22)=(13)/(17)	Depreciation cost support contract	5,2	1,9
Calculation	(23)=(14)/(17)	Depreciation cost auxiliary equipment	15,1	9,5
Calculation	(24)=(18)+(19)+(20)+(21)+(22)+(23)	**Total CAPEX**	**109,1**	**63,0**
Input	(25)	%Software cost over investment	4%	2%
Calculation	(26)=(25)*(8)	Software cost	16,49	3,55
Ratio	(27)	FTEs/1000 CPUs	4,61	2,16
Input	(28)	Average cost per FTE	54.207	48.800
Calculation	(29)=(27)*(28)/1000	Personnel cost / core	249,8	105,5
Ratio	(30)	m2/'1000 core	64,96	52,41
Assumption[70]	(31)	Average cost per m2	100	100
Calculation	(32)=(30)*(31)/1000	Site premises cost	6,496	5,241
Ratio	(33)	Electricity consumption per core in kWh	339,15	307,74
Assumption[71]	(34)	Cost /kwh	0,1	0,1
Calculation	(35)=(33)*(34)	Electricity cost	33,915	30,774
Input	(36)	Connectivity costs	0	0
Input	(37)	Other costs	0	0
Calculation	(38)=(26)+(29)+(32)+(35)+(36)+(37)	**Total OPEX**	**306,7**	**145,0**
Calculation	(39) = (24)+(38)	**Total yearly cost**	**415,82**	**208,04**
Input	(40)	**Utilisation rate**	**65%**	**75%**
Calculation	(41) = (39)/[8760*(40)]	**Cost per core hour**	**0,0730**	**0,0317**

Types of data:

Input: Information that is reported in the same format as used in the calculations

[70] We have used as a reference prices indicated on http://www.globalpropertyguide.com/faq/guide-sqm-prices-rents-yields. Nevertheless, the €100/m2 per year is a rough estimation.

[71] http://www.energy.eu/#industrial End-user energy prices for industrial consumers reference November 2011

Ratio: Information generated by combining more than one pieces of information presented in the questionnaire (e.g. FTEs per 1,000 cores; this ratio is created by dividing the number of FTEs by the number of cores reported by the respondent and by then multiplying with 1,000).
Assumptions: Assumptions used in the development of the model (e.g. the cost per kwh = €0.10).
Calculations: This is self explanatory.

7.4 Sample characteristics and descriptive statistics (appendix for 4.2)

This appendix presents analytical information about the characteristics of the respondents. The numerical data refer only to year 2011. Information regarding 2010 is available in Del. 2.2. "Computing e-Infrastructure cost calculation at national and European level" (http://www.efiscal.eu/deliverables).

Type of services provided by respondents

Our final sample consisted of 28 respondents. 14 out of the 28 respondents provide both computing and coordination activities while 13 provide only computing services. Only one respondent was solely a coordination organization.

What type of services does your institute provide?		
Answer Options	**Response Percent**	**Response Count**
Coordination (no resources)	4%	1
Computing services (CPU, storage, etc.)	46%	13
Both	50,0%	14
answered question		**28**

Table 14: Types of services represented in e-FISCAL survey

Part of e-Infrastructure in which respondents participate

Out of the 28 respondents, 14 (50%) indicated that their institute is only part of NGI/EGI, 3 reported participating into National HPC infrastructure/PRACE while another 8 participate in both (see Table 16).

To which e-Infrastructure is your institute's infrastructure part of?		
Answer Options	**Response Percent**	**Response Count**
NGI/EGI	50%	14
National HPC infrastructure/PRACE	10,7%	3
Both	28,6%	8
Other (please specify)*	17,9%	5
answered question		**28**

* There are two questionnaires indicating that the sites are "National HPC infrastructure/PRACE and other" and "Both and other"

Table 15: Institute's infrastructure participation represented in e-FISCAL survey

Types of institutes participating

The majority of institutes participating in e-FISCAL survey (20 out of 28) are NGI Resource Centres (CPU, storage, etc.), while another 15 out of 28 are HPC centres. An analytical review of these combinations is presented in Table 5.

Type of institute	Total
1,2	1
1,2,3	1
1,2,3,4	1
1,2,4	4
1,2,3,4,5	1
1,3	1
2	7
2,3,4	1
2,4	2
2,4,5	1
2,5	1
3,4	1
4	3
4,5	1
5	2
Total	28

Where:

1= NGI (coordinating body), 2= NGI Resource Centre (CPU, storage, etc.), 3= PRACE country coordinator, 4= HPC centre and 5 = Other (please specify)

Table 16: Type of Institutes participating in e-FISCAL survey

Hardware information

As for the e-Infrastructure of the respondents information about CPU cores and storage information follows

The CPU cores information of the sample is presented in Table 17.

Please present information in relation to the total number of "logical" CPUs (i.e. number of processing cores) of the NGI site/ HPC Centre available at the end of year 2011.				
Answer Options	**Min**	**Max**	**Average**	**Median**
Logical CPUs as on 31/12/2011	72	17,355	5,184	3,305

Table 17: CPU cores information

Information about the Storage capacity of the sample is presented in Table 18.

Please present information in relation to the TB of storage devices available in the NGI site/ HPC Centre at the end of year 2011.				
Answer Options	**Min**	**Max**	**Average**	**Median**
Disk Storage in TB as on 31/12/2011	3	13,070	1,207	333
Tape Storage in TB as on 31/12/2011[72]	12	7,350	1,903	980

Table 18: Storage information

Information about the acquisition costs of both hardware and storage of the sample is presented in Table 19.

Please present the average acquisition (i.e. purchase) cost per logical CPU and the average cost per TB acquisition in 2011. In case you have no data for 2011 please use approximations based on the most recent procurements or budget data. Note: Please do not include any hardware support contract costs or software costs in the values presented below				
Answer Options	**Min**	**Max**	**Average**	**Median**
Cost per logical CPU in € in 2011	80	800	275	218
Cost per TB/ Tapes in € in 2011	37	170	88	78
Cost per TB/ Disks in € in 2011	80	3,000	493	270

Table 19: Acquisition cost of logical CPU and storage information

The cost per logical CPU and the cost per TB exhibit significant differences. As one respondent commented these costs are rather dependent on the choice of compute/storage technology. It can be easily seen that the number of answers is this question are less that the total sample. According to informal discussions the disclosure of procurement information related to prices is considered sensitive data and therefore some respondents refrained from presenting their case[73]. The limited number of responses in relation to cost per TB for tapes is probably related to the fact that not all respondents use tapes as a storage medium.

Depreciation rates

The average useful lives for hardware and software are presented in Table 20. Depreciation rate can be calculated accordingly.

[72] We have included in descriptive only sites have non zero tape storage.

[73] We expanded our information set regarding hardware (computing and storage) procurement prices by getting information from the EGI compendium study. This type of information was used for cross-checking purposes. The median cost per logical CPU in EGI compendium was € 192 (6 responses) while the cost per TB € 373.

Please indicate the period in number of years that corresponds to the average useful economic life (depreciation period) of the following assets according to the policy followed by the NGI site/ HPC Centre.				
Answer Options	**Min**	**Max**	**Average**	**Median**
Average useful life in years for CPUs	3	10	4.9	5
Average useful life in years for tape storage devices	3	12	6.5	5
Average useful life in years for disk storage devices	3	20	5.6	5

Table 20: Useful life information

Analysing Table 20 allows us to make an interesting observation. Typically the Total Cost of Ownership (TCO) exercises use primarily a four year duration (e.g. Walker (2009) and Magellan report (2011) and sometimes a three year duration (e.g. Nazir and Sørensen (2010)) – as the useful lifetime of a CPU. In our sample the median useful life to calculate annual depreciation is five years. This has a non-negligible impact on the capital costs accounted each year in order to form the total yearly cost. It also has significant effects on electricity consumption as older machines consume more electricity. Nevertheless, there are instances of long depreciation periods which will require deeper study to understand the underlying conditions that permit such a treatment.

Cost relations over investment

In order to assess several other cost parameters that are related to CAPEX but at the same time are not easily identifiable and measurable, we asked respondents to provide indications on their relative size compared to the computing and storage investment. As these estimations were expected to be rough, respondents were given the flexibility instead of identifying a sole number (in our case percentage) to indicate a range of values (e.g. between 10%-20%). Therefore the following numbers are considered as high-level approximations.

Estimated cost relations of several parameters on computing and hardware storage				
	Min	**Max**	**Average**	**Median**
Please present an overall estimation of the related interconnect equipment costs (network devices, cables, etc.) as a percentage of the hardware acquisition cost*	0%	30%	10%	9%
Please present an overall estimation of the support contract costs (e.g. next-business-day hardware support costs) as a percentage of the hardware (CPUs and storage devices) acquisition cost **	0%	25%	6%	4%
If you were to equip the existing NGI site/ HPC Centre now what would be the investment cost of all auxiliary equipment as percentage of the cost of acquiring computing and hardware storage capacity***	5%	35%	18%	20%
Please make an estimation of the total cost of the related software (e.g. operating system, fabric layer / file system software (e.g. LSF, GPFS), software support contract costs, applications cost, 3rd party software cost, compilers, etc.) as a percentage of the hardware (CPUs and storage devices) acquisition cost	0%	15%	4%	2%

**A respondent commented that for CPU, first level switches are included in the servers' price. For instance, blade centres already include switches.*
*** Four respondents made comments on the "other" option in this question. Their comments could be summarised as follows: These support terms are included in hardware tender specifications (for 3 years or 3-5 years) and are difficult to separate (not typically included in contract as a line item). Apart from that some centres pay extra money for support while others don't.*
**** Due to the wording of the question the analyses that follows has used 50% of the average percentage value of auxiliary equipment over investment in order to calculate the yearly cost per core and cost per core hour.*

Table 21: Cost relations over investment

The software costs are most accurately characterised as an operating expense in today's e-Infrastructure ecosystem. The majority of the system software used today is produced in collaborations (in which many of the service providers participate) and released under open source license, which can explain both the relatively low proportion of the costs and large variations between centres.
This contrast with the situation in many of the ICT environments outside the research domain, where the software licenses for applications such as enterprise resource planning (ERP) systems may form the biggest part of the capital expenditure, and a division of the software costs between CAPEX (license fees) and OPEX (service contracts) is necessary[74]. In Intersect 360 report a significant difference between software expenditure for academia and business is witnessed; academia spending significantly less on software licences and maintenance costs.

Personnel information

Salary levels exhibit significant differences among countries (ref. Min and Max values in Table 22). It could be argued that personnel cost is a cost element that is highly dependent on local - country conditions; a phenomenon which is less evident as far as technology, type of infrastructure or architecture are concerned (i.e. the procurement of hardware or storage in two different counties is not expected to deviate to such an extend only due to the geographical area). Therefore the structure of our sample (i.e. counties participating in the survey) form some metrics of average and median personnel costs that for some other countries could be exceptionally high or low respectively. Interestingly enough the median cost per FTE from the EGI compendium data set is €44.000 while the average cost is €46,200 close to our findings.

Please provide the following information related to the average yearly salary per FTE.				
Answer Options	**Min**	**Max**	**Average**	**Median**
Average yearly salary cost per FTE (gross salary plus employee benefits and bonuses) **in '000** € in 2011	15	108	54.21	48.80

[74] It should be noted that assessing the overall costs of the European e-Science ecosystem are outside of the scope of this study. However, it is possible to assume that algorithmic developments and software maintenance tasks that are directly related to research would require similar amounts of effort in dedicated or on-demand infrastructure. The overall structure and behaviour of the scientific software ecosystem is a topic of an e-IRG Task Force on Scientific Software.

Table 22: Salary information

Space and PUE information

In the following table information about the site space and PUE is presented. We acknowledge that in several cases this type of information is not systematically kept. However, due to the increasing awareness on Green IT aspects these data would gain more and more importance especially due to the increased cost of electricity consumption.

Please fill in the following information related to the cost and operating characteristics of the NGI site/ HPC Centre for 2011.				
Answer Options	**Min**	**Max**	**Average**	**Median**
Site centre space in m2 (2011)	6	1,000	228.25	133
Power Usage Effectiveness in 2011	1.25	2.24	1.56	1.50

Table 23: Site information

The questions about network connectivity costs and questions relating to overhead[75] costs were only sparsely competed. Moreover, when these pieces of information were reported the magnitude of costs for both connectivity and other overhead was in general small and it looks like, at least the e-FISCAL sample, that it would not make a difference in the calculated amounts. Therefore these cost categories were not subsequently analysed. As there is not enough information we do not report any values about these two cost categories. However, we keep these cost categories as part of the overall cost breakdown with zero amounts. Notwithstanding that their non-inclusion is not expected to have a material effect, we accept that by taking into consideration such costs, cost calculations would have been driven, marginally, up.

Cloud computing and sustainability considerations

The last part of the questionnaire contained questions regarding cloud computing, Green IT aspects and sustainability considerations. The time horizon of these questions tried to balance recent past actions (i.e. actions that have been realized in 2011) and future prospects.

As for cloud computing, it is apparent from Table 24 the most frequently used cloud computing service is Software as a Service (SaaS), which counts 10 answers (8 respondents used will 2 bough such services) followed by Infrastructure as a Service (IaaS) with 7 answers.

Please answer the following questions in relation to the use of cloud computing in 2011: Note: "Use" means free of charge		
Answer Options	**Use**	**Buy**
Did you use (buy) Infrastructure as a Service (e.g. Amazon EC2) in 2011?	4	3
Did you use (buy) Platform as a Service (e.g. Microsoft Azure) in 2011?	2	0
Did you use (buy) Software as a Service (e.g. Google Docs, Microsoft Live	8	2

[75] The overhead costs refer to travelling expenses, conferences, training and insurance.

services) in 2011?		
Did you use (buy) disk storage services from external providers in 2011?	2	0
Did you use (buy) tape storage services from external providers in 2011?	0	0

Table 24: Using and buying cloud computing services in 2011

However the prospects for the future are in favour of a more intensive, compared to 2011, cloud computing services usage (Table 25). Nevertheless, most respondents are more prone to use (free of charge) cloud computing services compared to those expressing their intention to buy such services.

Please answer the following questions in relation to the use of cloud computing in the future:		
Answer Options	**Use**	**Buy**
Do you intend to use (buy) Infrastructure as a Service in the future?	8	5
Do you intend to use (buy) Platform as a Service in the future?	6	0
Do you intend to use (buy) Software as a Service in the future?	11	3
Do you intend to use (buy) disk storage services from external providers in the future?	9	1
Do you intend to use (buy) tape storage services from external providers in the future?	3	0

Table 25: Using and buying cloud computing services from 2012 onwards

Moreover, one respondent raised some concerns regarding cloud computing that refer to three dimensions: a) the cost of using cloud services to transfer and store large amounts of data (e.g. hundreds of TB), b) security issues related to the safeguarding of data confidentiality (i.e. confidential data stored to third party storage devices), and c) CPU capacity (that is translated into additional costs) necessary to encrypt and regularly decrypt the data stored in third party premises.

Finally as it is apparent in the following table the idea of imposing usage fees is not out of the question for several respondents that have included such provision in both the short and long term planning.

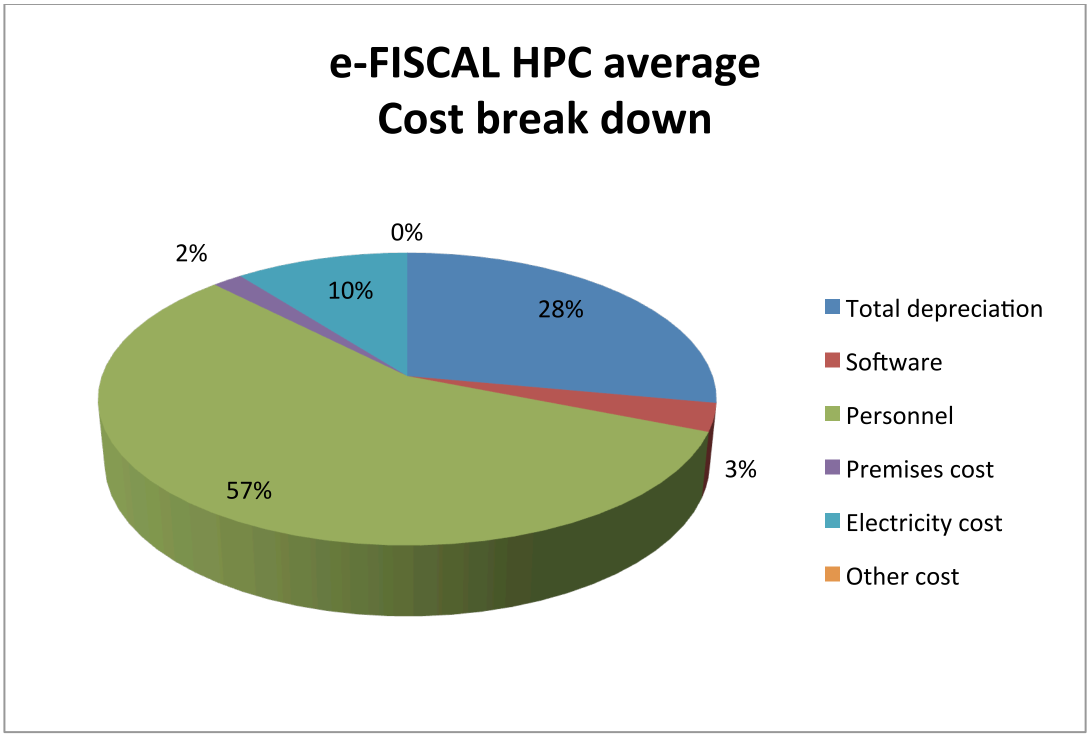
Please answer the following sustainability related questions:		
Answer Options	**Yes**	**No**
Do you have a short-term (e.g. 1–3 years) capacity and business plan for your computing infrastructure?	21	2
Do you have a long-term (e.g. 3–5 years) capacity and business plan for your computing infrastructure?	13	9
Is there a provision of any kind of usage fees in the short-term plan?	10	13
Is there a provision of any kind of usage fees in the long-term plan?	9	14

Table 26: Sustainability related information

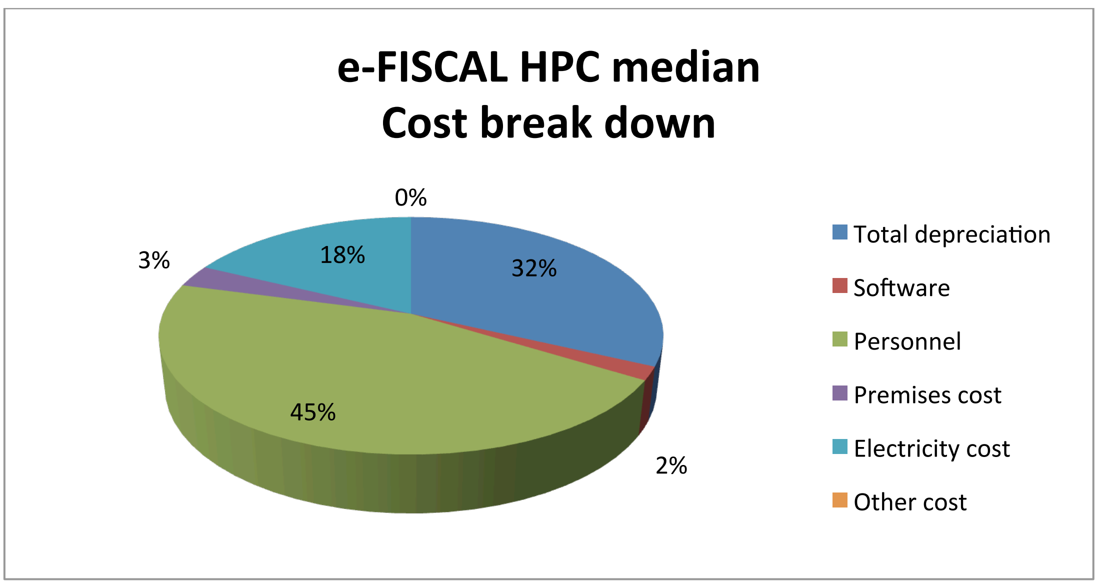
Information about the cost break down of e-FISCAL HPC centres – Base Case split

The following graphs present the cost break down under the costing Option 2.

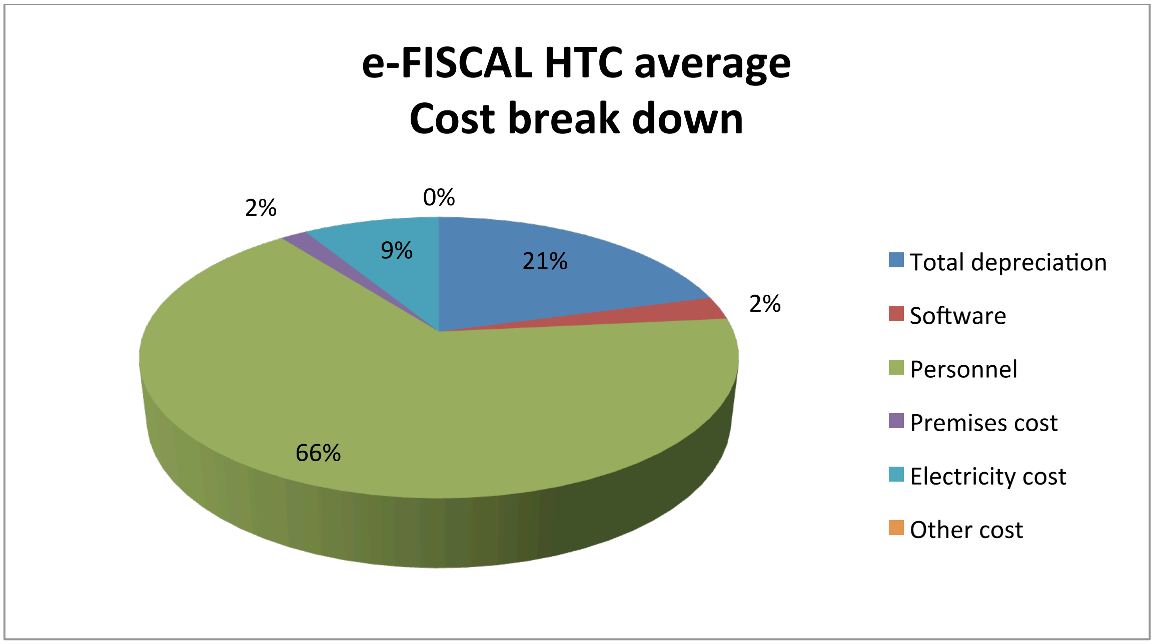
The total depreciation refers to the computing depreciation and actually corresponds to yearly CAPEX.



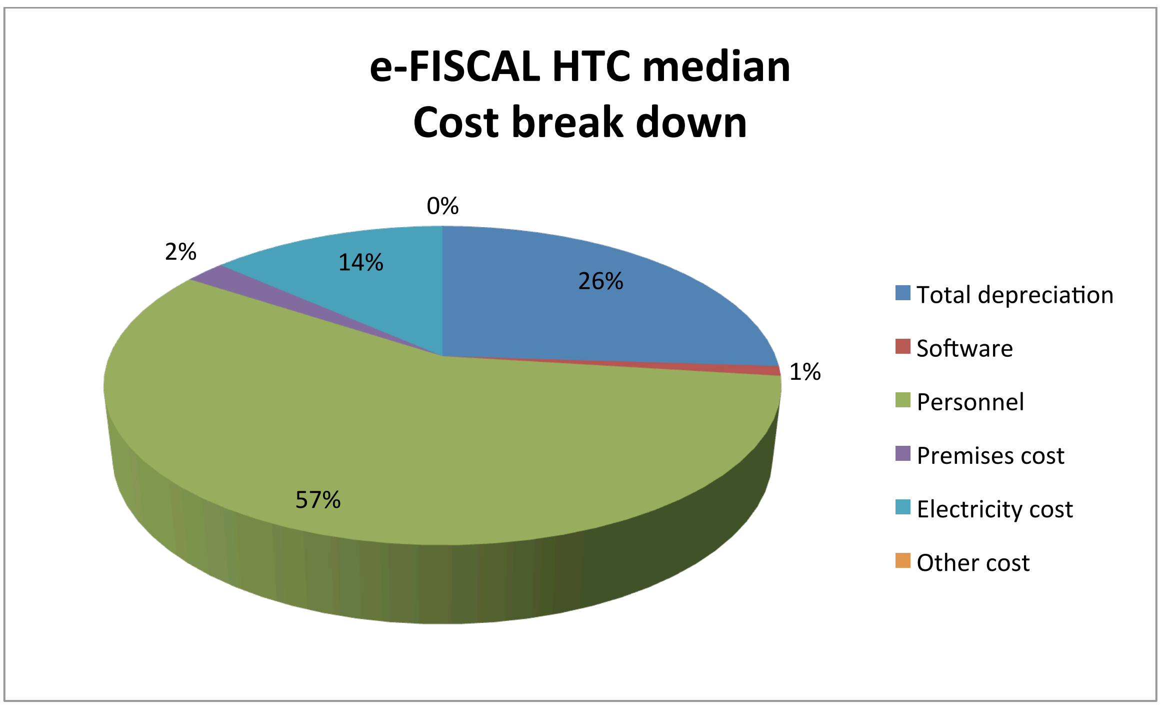
Graph 15: e-FISCAL HPC cost break down (average)



Graph 16: e-FISCAL HPC cost break down (median)



Graph 17: e-FISCAL HTC cost break down (average)



Graph 18: e-FISCAL HTC cost break down (median)

7.5 Results of Univariate OLS regressions

The following table shows the results of the univariate Ordinary Least Square regressions. The only statistically significant conclusion (at 10% level of statistical significance) that can be drawn from the table below is that the number of FTEs /1,000 cores is smaller the larger the site. Stata 12 was used to run the statistics.

VARIABLES	(1) Cost per core	(2) Cost per core hour	(3) FTEs/1,000 cores
Cores/1,000	-4.407	-0.001	-0.375*
	[8.534]	[0.001]	[0.206]
Constant	414.347***	0.067***	6.412***
	[62.563]	[0.010]	[1.514]
Observations	28	28	28
R-squared	0.010	0.020	0.112
Adj R-Squared	-0.0279	-0.0173	0.0783

Standard errors in brackets
*** p<0.01, ** p<0.05, * p<0.1

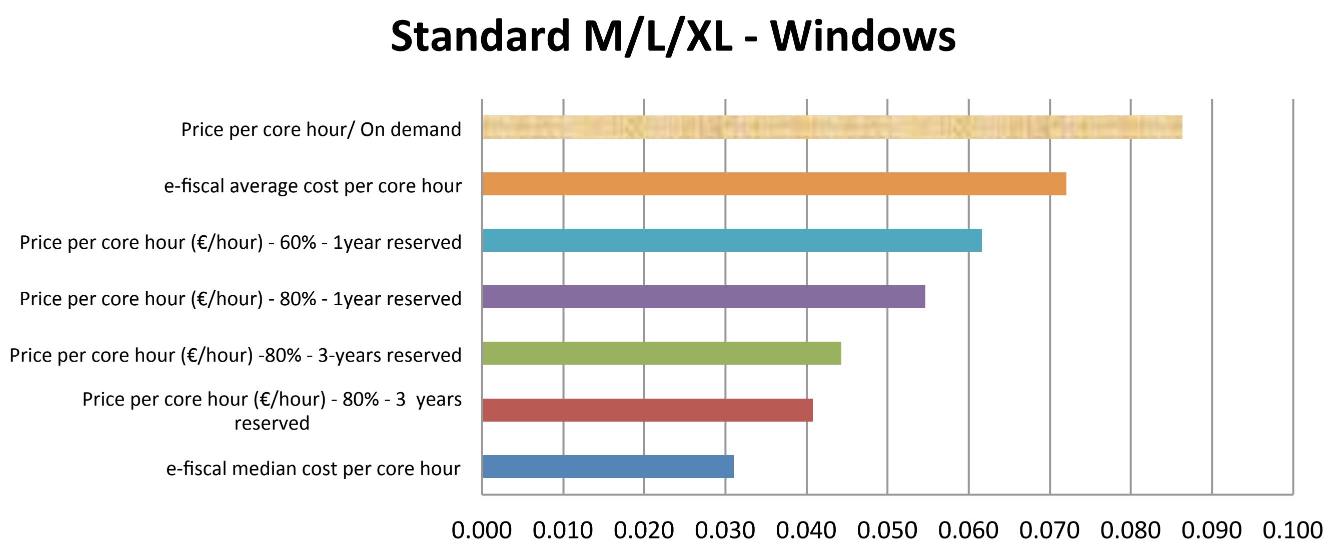
Table 27: Univariate OLS regression results

Univariate OLS regression models:

(1) Cost per core = constant + b x cores/1,000 + e
(2) Cost per core hour = constant + b x cores/1,000 + e
(3) FTEs/1,000 = constant + b x cores/1,000 + e

7.6 Price comparisons (Amazon Instances - Windows)

The following graphs present the results of the comparison between e-FISCAL costs and Amazon Windows Instances.

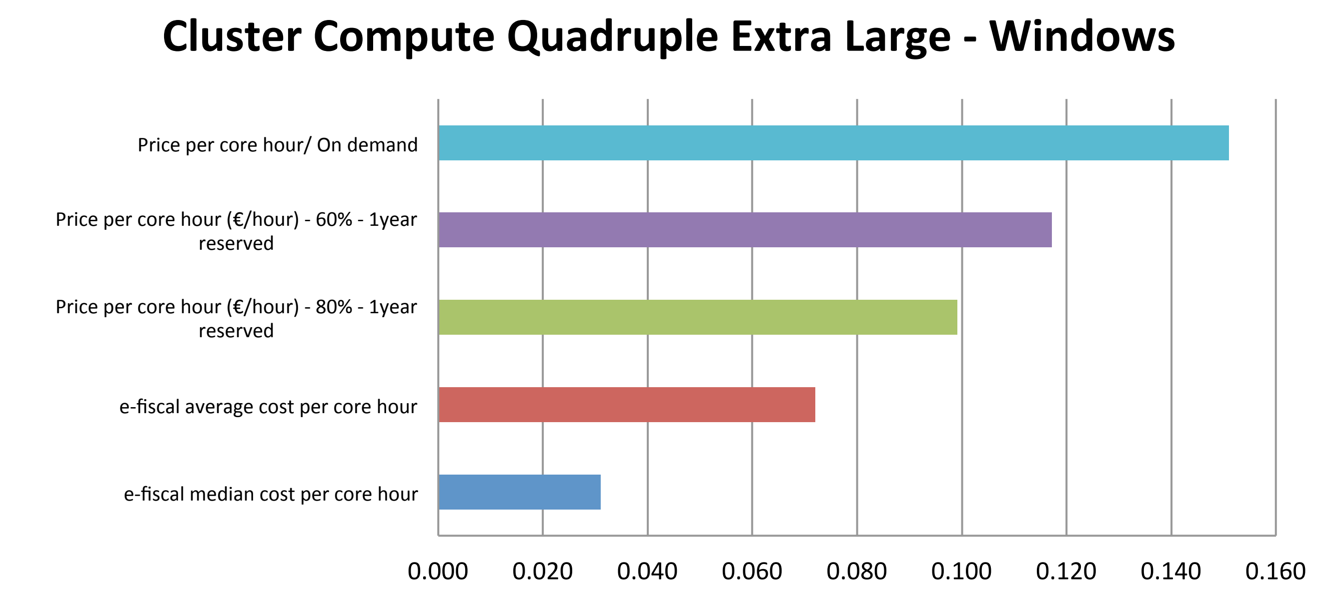


	e-fiscal median cost per core hour	Price per core hour (€/hour) - 80% - 3 years reserved	Price per core hour (€/hour) -80% - 3-years reserved	Price per core hour (€/hour) - 80% - 1year reserved	Price per core hour (€/hour) - 60% - 1year reserved	e-fiscal average cost per core hour	Price per core hour/ On demand
Standard M/L/XL - Windows	0.031	0.041	0.044	0.055	0.062	0.072	0.086

Graph 19: Comparing e-FISCAL cost with Amazon prices (Standard M/L/XL instances – Windows)

By taking into account the Amazon prices the median e-FISCAL cost per core hour (basic case) is less expensive even than the 3-year reserved instances. Therefore it is less expensive in all cases.

The average e-FISCAL cost per core hour (basic case) is more expensive compared to the prices of the reserved instances but less expensive than the on-demand instances.



	e-fiscal median cost per core hour	e-fiscal average cost per core hour	Price per core hour (€/hour) - 80% - 1year reserved	Price per core hour (€/hour) - 60% - 1year reserved	Price per core hour/ On demand
Cluster Compute Quadruple Extra Large - Windows	0.031	0.072	0.099	0.117	0.151

Graph 20: Comparing e-FISCAL cost with Amazon prices (Cluster Compute Quadruple XL instances – Windows)

The median e-FISCAL cost and the average e-FISCAL cost (basic case) per core hour are less expensive than the prices of both on demand and one-year reserved instances.

7.7 Benchmarking exercise results

HTC Benchmarking – System Specifications

This section will cover the details about the HTC benchmarking setup and system specifications used to compare the in-house vs. cloud computing instances. An introduction to the HEPSEPC06 benchmark that is used for the HTC benchmarking is already included in the deliverable D2.2. Moreover, for HPC benchmarking system specifications and benchmark results, please also refer to the D2.2.

In-House HTC Systems Specification

Two different systems are used in-house at the INFN data centre to run the HEPSEPC06 benchmark. Below are the system specifications for the two systems:

- **System with Intel Sandy Bridge**
  - **Hardware:** Intel(R) Xeon(R) CPU E5-2660 @ 2.2 GHz, 2 x 8 cores, 64 GB memory
  - **Software:** Scientific Linux 6.2, 64-bit platform with Hyper-Threading enabled. GNU C compiler.
- **AMD Opteron**
  - **Hardware:** 6272 (aka Interlagos) @ 2.1 GHz, 2 x 16 cores, 64 GB memory
  - **Software**: Scientific Linux 6.2, 64-bit platform, GNU C compiler.

Amazon EC2 Instances Specification

Following Amazon EC2 instances types are used to run the HEPSPEC06 benchmark: M1 Medium, M1 Large, M1 Extra-Large, M3 Extra-Large and M3 Double Extra-Large. Below are the system specifications for each of the instance:

- **M1 Medium Instance**
  - **Hardware:** 2 EC2 Compute Unit (1 virtual core with 2 EC2 Compute Unit), 410 GB instance storage, 3.75 GB memory
  - **Software:** Scientific Linux 6.2, 64-bit platform with Hyper-Threading enabled. GNU C compiler.
- **M1 Large Instance**
  - **Hardware:** 4 EC2 Compute Units (2 virtual cores with 2 EC2 Compute Units each) 850 GB instance storage, 7.5 GB memory
  - **Software:** Scientific Linux 6.2, 64-bit platform with Hyper-Threading enabled. GNU C compiler.
- **M1 Extra-Large Instance**
  - **Hardware:** 8 EC2 Compute Units (4 virtual cores with 2 EC2 Compute Units each) 1,690 GB instance storage, 15 GB memory
  - **Software:** Scientific Linux 6.2, 64-bit platform with Hyper-Threading enabled. GNU C compiler.
- **M3 Extra-Large Instance**
  - **Hardware:** 13 EC2 Compute Units (4 virtual cores with 3.25 EC2 Compute Units each), 15 GB memory
  - **Software:** Scientific Linux 6.2, 64-bit platform with Hyper-Threading enabled. GNU C compiler.
- **M3 Double Extra-Large Instance**
  - **Hardware:** 26 EC2 Compute Units (8 virtual cores with 3.25 EC2 Compute Units each), 30 GB memory
  - **Software:** Scientific Linux 6.2, 64-bit platform with Hyper-Threading enabled. GNU C compiler.

The following Table 28 summarises the system specification for both In-house HTC and Amazon EC2 Cloud instances:

	Amazon EC2	HTC resource at INFN
Compute Nodes	**Medium: 2 ECU** **Large: 4 ECU** **XL: 8 ECU** **M3 XL: 13 ECU** **M3 2 XL: 26 ECU** ***1 ECU = 1.0-1.2 GHz of*** **Intel Xeon or AMD Opteron**	**- Intel(R) Xeon(R) CPU E5-2660 @ 2.2 GHz, 2 X 8 cores, 64 GB memory** **- AMD Opteron 6272 (aka Interlagos) @ 2.1 GHz, 2 X 16 cores** • ***M instance Single-core VM*** • ***L instance Dual-core VM*** • ***XL Instance Quad-core VM*** • ***M3 XL Instance Quad-core VM*** • ***M3 Double XL Instance Eight Core VM***
OS	**SL6.2, 64-bit platform**	**SL6.2, 64-bit platform**
Memory	**3.75 GB, 7.5 GB and 15 GB**	**64 GB for both Intel and AMD**
Hyper-Threading	**Enabled**	**Enabled (for Intel) ~ 32 logical cores**
Compilers	**GCC**	**GCC**

Table 28 – Benchmarking system specification

HTC Benchmarking – Settings and Results

This section will cover the details on the benchmarking settings that are deployed to compare the in-house and Amazon EC2 instances. As a first step, the following sub-section will detail the various benchmark settings.

Benchmark Settings

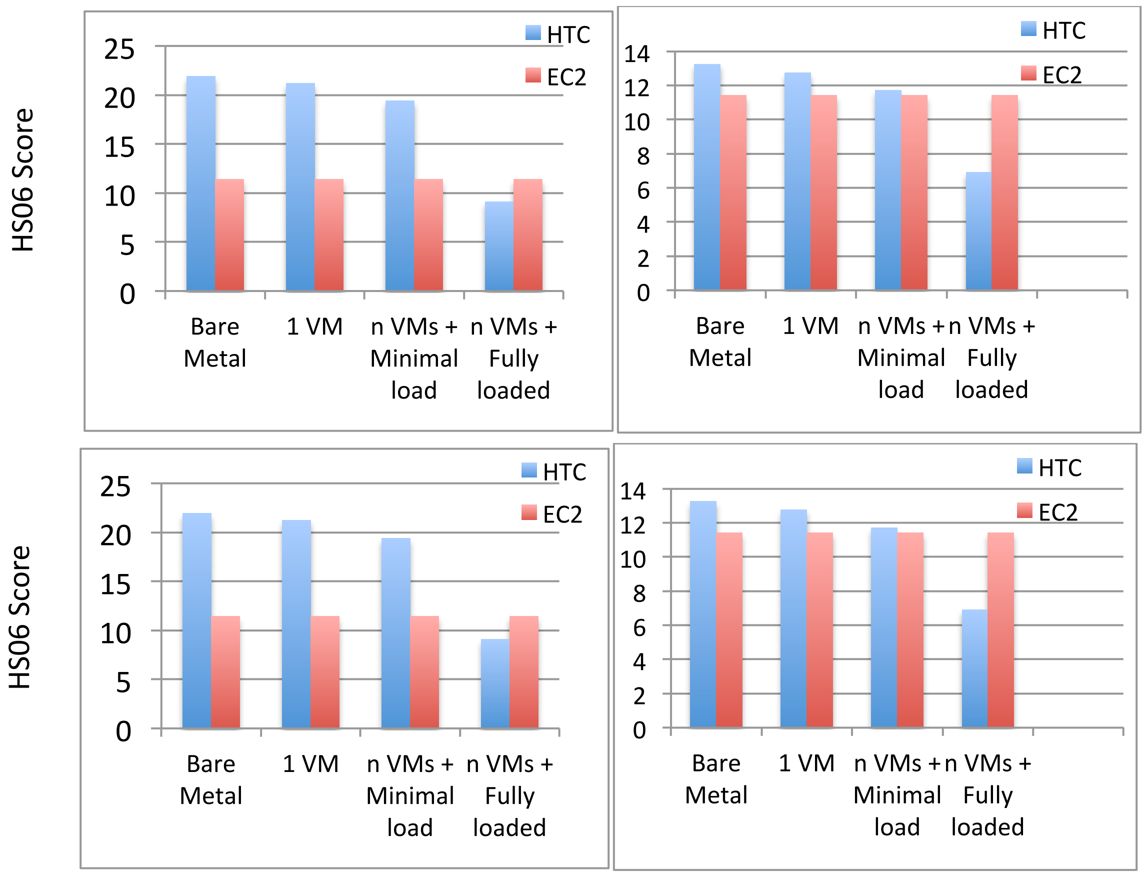
The rationale behind the following benchmark settings is to try to simulate the public Cloud characteristics i.e. virtualization, multi-tenancy, resource over-subscription and load variations. As there is no other simple way to study the effect of these distinct Cloud characteristics within the Cloud computing infrastructures. The following benchmark settings are adopted:

- **HEPSPEC6 on Bare-metal vs. EC2:** The HEPSPEC06 results on bare-metal vs. EC2 demonstrates the performance impact while running applications on non-virtualised infrastructure vs. EC2 virtualised and multi-tenant infrastructure.
- **HEPSPEC6 on 1 virtual machine vs. EC2:** The HEPSPEC06 result on 1 virtual machine compared against the bare-metal case yields the virtualization over-head. However, it is unlikely to have a comparable case in a commercial cloud setting where there will be only one virtual machine running in the whole infrastructure.
- **HEPSPEC6 on n virtual machines with minimal workload vs. EC2:** In order to simulate the effect of multi-tenancy physical machine was loaded with the maximum number of possible virtual machines it could accommodate without any resource oversubscription. And then the HEPSPEC06 results are compared for both in-house and EC2 Cloud instances. This test case simulates the effect of multi-tenancy with minimal workload.
- **HEPSPEC6 on n virtual machines with maximum workload vs. EC2:** To simulate the effect of worst-case scenario, all machines within the multi-tenant in-house infrastructure are fully loaded with HEPSEPC06 and then results are compared with the EC2 instances.

Benchmark Results and Discussion

In this Section, results from the various benchmark runs are included along with the highlights of the interesting observations.

HS06 for M1 Medium instance – Intel **HS06 for M1 Medium instance - AMD**



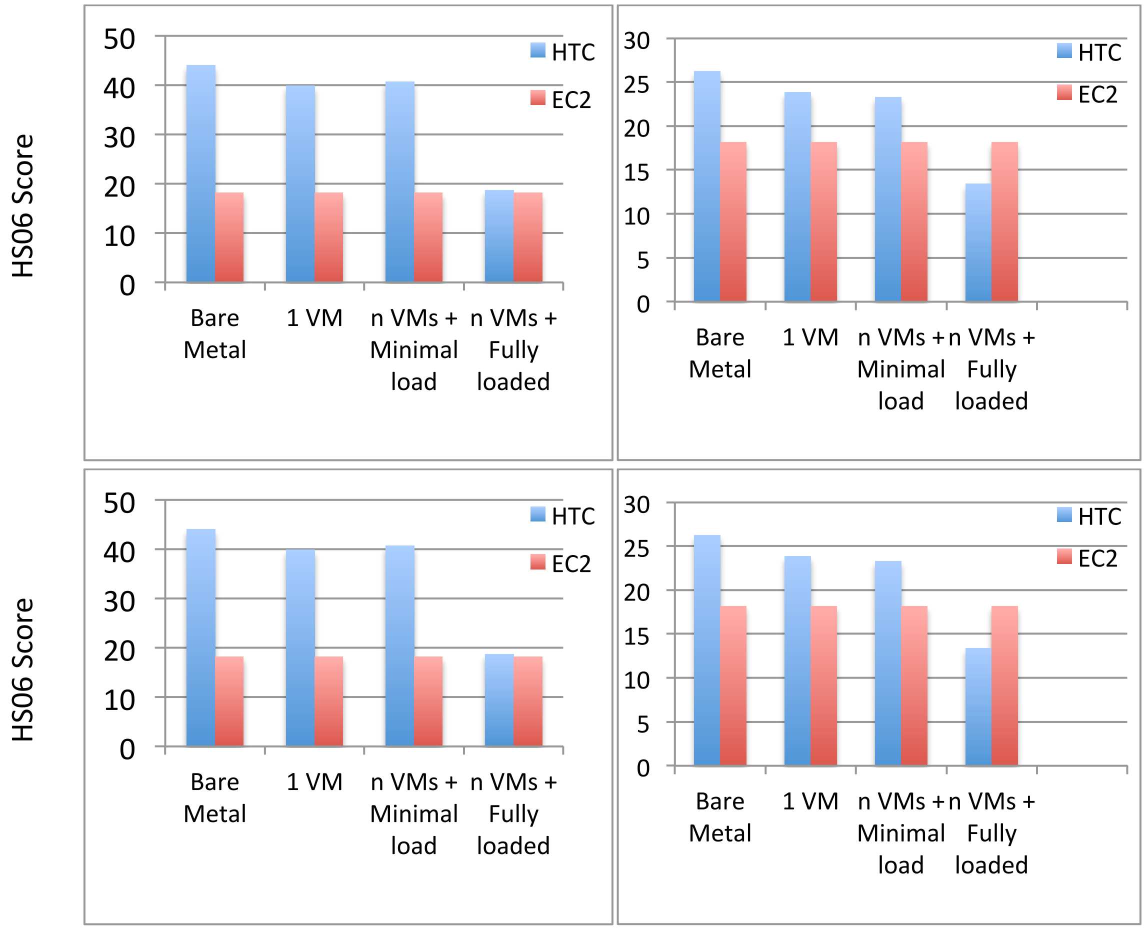
HTC Number of VMs = 32 **HTC Number of VMs = 32**

The two graphs above depict the HEPSEPC06 scores comparison for the EC2 M1 Medium instance and corresponding in-house HTC instance at INFN with Intel and AMD processor. From the graphs following observations could be made:

- Intel processor is better in performance as compared to the AMD Opteron (please note the y-axis scale for Intel vs. AMD).
- Performance degradation while comparing the AMD Opteron bare-metal and EC2 virtualised instance are significantly low as compared to the Intel Sandy Bridge.
- For the worst-case (i.e. all VMs running HEPSEPC06) the EC2 outperforms both the in-house instances i.e. with Intel and AMD processors. In all the other cases, in-house HTC instance is better as compared to the EC2.
- The minimal effect of virtualization is notable by comparing the HEPSPEC06 scores for the bare-metal vs. 1 virtual machine. Also, this trend continues even with the increase in the number of virtual machines.
- 47.95% average performance loss (ranging from 2.38-58.48%) while comparing the HEPSEPC06 scores between EC2 and HTC instances (with Intel processors) at INFN.

- 18.19% average performance loss (ranging from 3.77-47.89%) while comparing the HEPSEPC06 scores between EC2 and HTC instances (with AMD processors) at INFN.

HS06 for M1 Large instance – Intel **HS06 for M1 Large instance – Intel AMD**

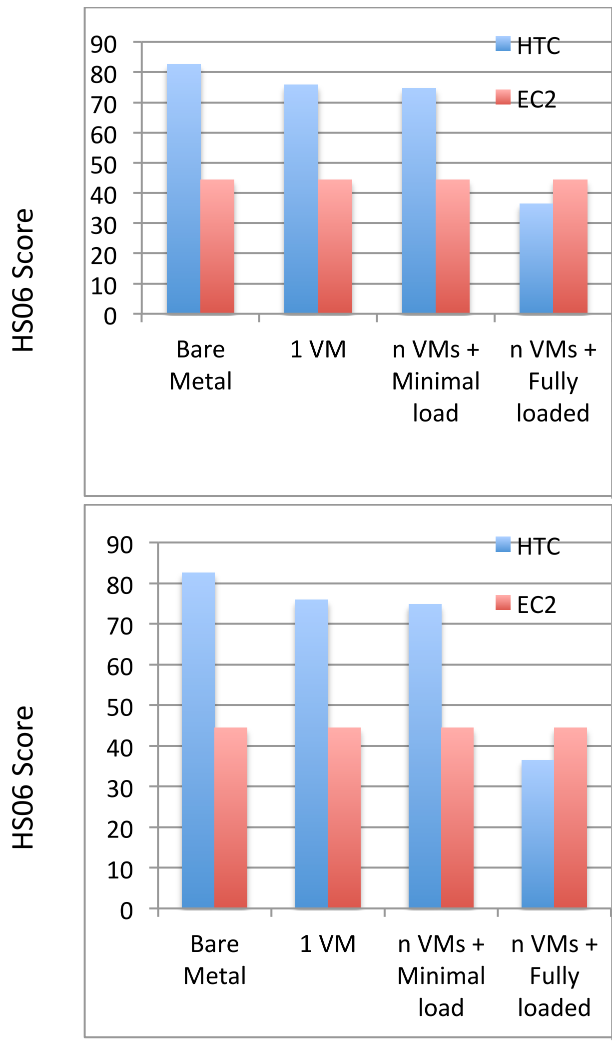


HTC Number of VMs = 16 **HTC Number of VMs = 16**

The two graphs above depict the HEPSEPC06 scores comparison for the EC2 M1 Large instance and corresponding in-house HTC instance at INFN with Intel and AMD processors. From the graphs following observations could be made (in addition to what has already been discussed):

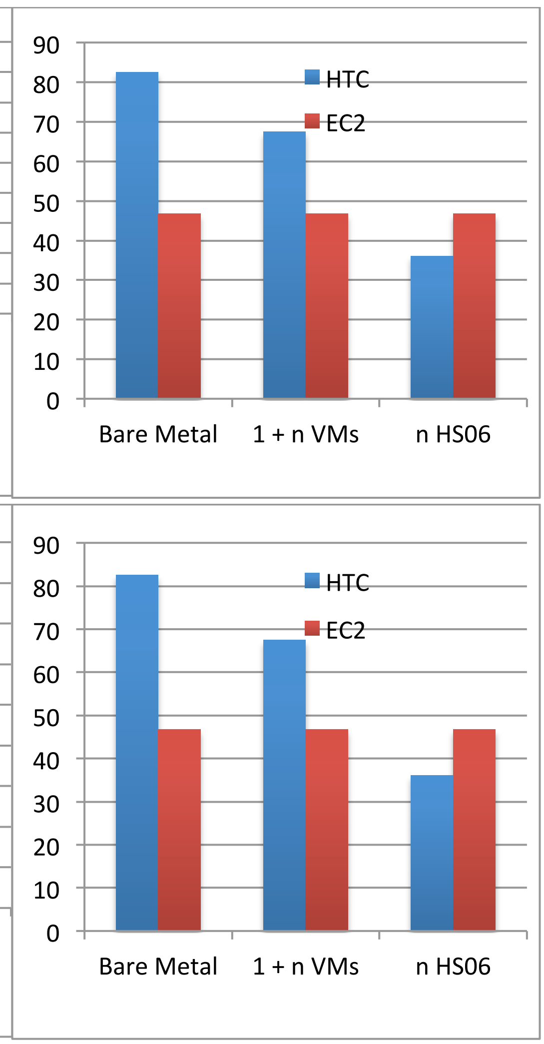
- Large instance has attained better HEPSPEC06 results, thus exhibit better performance as compared to the medium instance due to better specifications.
- For the worst-case (i.e. all VMs running HEPSEPC06) both the EC2 and in-house instances (i.e. with Intel processors) achieved more or less similar scores. This could lead to the conclusion that workloads were towards the upper bound, when benchmarks were executed within the EC2 infrastructure.
- 58.79% average performance loss (ranging from 9.49-57.47%) while comparing the HEPSEPC06 scores between EC2 and HTC instances (with Intel processors) at INFN.
- 30.75% average performance loss (ranging from 9.04-48.88%) while comparing the HEPSEPC06 scores between EC2 and HTC instances (with AMD processors) at INFN.

HS06 for M1 XL – Intel



HTC Number of VMs = 8

HS06 for M3 XL – Intel

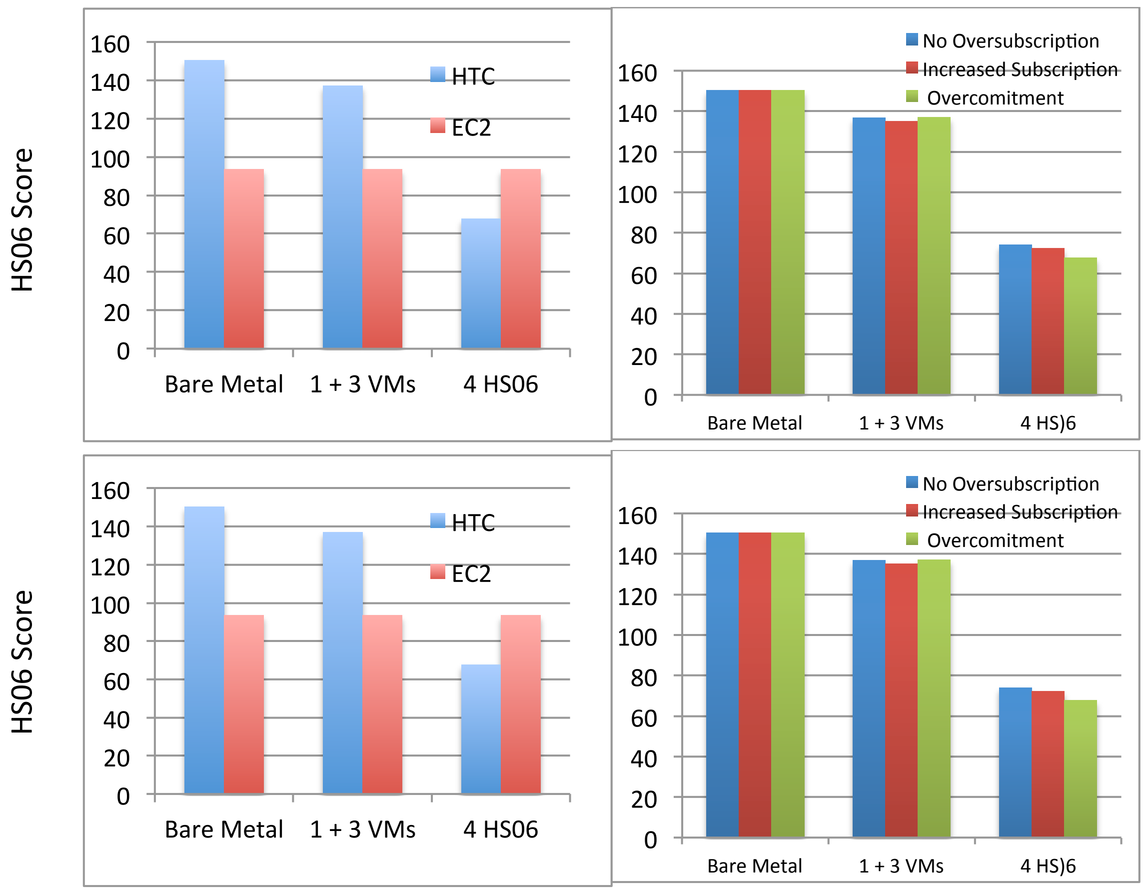


HTC Number of VMs = 8

The above two graphs illustrate the HEPSEPC06 scores comparison for the EC2 M1 and M3 Extra Large instances and corresponding in-house HTC instances at INFN with the Intel processors. From the graphs following observations could be made (in addition to what has already been discussed):

- 46.21% average performance loss (ranging from 8.14-55.84%) while comparing the HEPSEPC06 scores between M1 Extra-Large EC2 and corresponding in-house HTC instances at INFN.
- 43.37% average performance loss (ranging from 18.24-55.26%) while comparing the HEPSEPC06 scores between M3 Extra-Large EC2 and corresponding in-house HTC instances at INFN.

(a) HS06 for M3 Double XL – Intel **(b) Resource subscription impact**



HTC Number of VMs = 4 **HTC Number of VMs = 4**

The above graph (a) demonstrates the HEPSEPC06 scores comparison for the EC2 M3 Double Extra Large instances and corresponding in-house HTC instances at INFN with the Intel processors. From the graphs following observations could be made (in addition to what has already been discussed):

- 37.79% average performance loss (ranging from 8.81-54.98%) while comparing the HEPSEPC06 scores between M3 Double Extra-Large EC2 and corresponding in-house HTC instances at INFN.

In addition, (b) displays the HEPSPEC score trends with the increasing level of resource subscription and over-subscription. In the bare-metal case there is no memory over-subscription, as memory is solely utilized by one instance. In the second case, memory resource is shared amongst 4 VMs, where only one is running the HEPSPEC06 benchmark and rest of the three virtual machines are idle. We can see that there is no significant effect of increasing the memory subscription as long as it stays below the available memory size. The third case where all four virtual machines are running HEPSEPC06 and memory is stressed across the physical machine, it is clear that oversubscription of only the memory resource decreases the HEPSEPC06 score (thus increases the performance loss). Although more detailed test case is required to fully understand the effect of resource oversubscription.