SEVENTH FRAMEWORK PROGRAMME
Research Infrastructures

INFRA-2010-2.3.1 – First Implementation Phase of the European High Performance Computing (HPC) service PRACE

PRACE-1IP

PRACE First Implementation Project

Grant Agreement Number: RI-261557

D2.4.3
Initial Impact Assessment of the Research Infrastructure

Final version

Version: 1.2
Author(s):   Carlos Merida-Campos, BSC; Pawlikowska Katarzyna, ETH Zurich
Date: 25.6.2012

Project and Deliverable Information Sheet

PRACE Project	Project Ref. №: RI-261557	
	Project Title: PRACE First Implementation Project	
	Project Web Site: http://www.prace-project.eu	
	Deliverable ID: < D2.4.3>	
	Deliverable Nature: <DOC_TYPE: Report>	
	Deliverable Level: PU *	Contractual Date of Delivery: 30 / June / 2012
		30 / June / 2012
	EC Project Officer: Thomas Reibe	

* The dissemination level are indicated as follows: **PU** – Public, **PP** – Restricted to other participants (including the Commission Services), **RE** – Restricted to a group specified by the consortium (including the Commission Services). **CO** – Confidential, only for members of the consortium (including the Commission Services)

Document Control Sheet

	Title: Initial Impact Assessment of the Research Infrastructure	
Document	ID: D2.4.3	
	Version: 1.2	Status: Final
	Available at: http://www.prace-project.eu	
	Software Tool: Microsoft Word 2007	
	File(s): D2.4.3.docx	
Authorship	Written by:	Carlos Merida-Campos, BSC Pawlikowska Katarzyna, ETH Zurich
	Contributors:	Sergio Bernardi, CINECA Georgi Prangov, NCSA Michael Browne, ICHEC Dominik Ulmer, ETH Zurich Michele de Lorenzi, ETH Zurich Manuel Fiolhais, UC-LCA Jussi Heikonen, CSC Pekka Manninen, CSC Marcin Ostasz, BSC Sjoerd Meihuizen, NCF Nicolas Megnerey, GENCI Nelly Stoyanova, NCSA Pascal Moussier, GENCI
	Reviewed by:	Jacko Koster (SIGMA), Florian Berberich (PMO)
	Approved by:	MB/TB

Document Status Sheet

Version	Date	Status	Comments
0.1	2/July/2011	Guidelines	Guidelines of content and initial materials to be further developed
0.2	29/March/2012	Structure with preliminary content	Distributed with the assigned editors and content identified
0.3	27/April/2012	First complete draft	With contributions from all participants except by NCF, CINECA and GENCI
0.4	27/May/2012	Refinement of the draft	With contributions from NCSA, UC-LCA, BSC, GENCI, CINECA
0.5	6/June/2012	Refinement of the draft	Edited by BSC
0.6	8/June/2012	Refinement of the draft	Semi-complete conclusions and input provided by ICHEC
0.7	9/June/2012	Refinement of the draft	Complete conclusions and input provided by ICHEC and NCSA
1.0	10/June/2012	Final	Refinements on industry related variables and Format check by CSCS
1.1	11/June/2012	Final	Added additional input by BSC, NCSA and ICHEC
1.2	25/June/2012	Final	Addressed comments received by the internal reviewers (Florian Berberick and Jacko Koster)

Document Keywords

Keywords:	PRACE, HPC, Research Infrastructure, Monitoring, Reporting, Impact Assessment

Table of Contents

List of Figures

List of Tables

References and Applicable Documents

[1]   Deliverable D2.4.1, Monitoring and Reporting Procedures, PRACE-1IP project.
[2]   Deliverable D2.4.2, Initial Impact Assessment of the Research Infrastructure, PRACE-1IP project.
[3]   Deliverable D3.2.1, Training Plan, PRACE-1IP project.
[4]   Deliverable D3.2.5 Final Training Report, PRACE-1IP project.
[5]   Deliverable D4.1 Training and Education Survey, PRACE-2IP project.
[6]   Deliverable D4.2 Establishment of PRACE Advance Training Centres, PRACE-2IP.

List of Acronyms and Abbreviations

AC	Access Committee
BoD	Board of Directors
BSC	Barcelona Supercomputing Center (Spain)
CINECA	Consorzio Interuniversitario, the largest Italian computing centre (Italy)
CINES	Centre Informatique National de l'Enseignement Supérieur (represented in PRACE by GENCI, France)
CPU	Central Processing Unit
CSC	Finnish IT Centre for Science (Finland)
CSCS	The Swiss National Supercomputing Centre (represented in PRACE by ETHZ, Switzerland)
EC	European Community
ETHZ	Eidgenössische Technische Hochschule Zuerich, ETH Zurich (Switzerland)
GENCI	Grand Equipement National de Calcul Intensif (France)
GPU	Graphic Processing Unit
HM	Hosting Member

HPC	High Performance Computing
ICHEC	Irish Centre for High-End Computing
JSC	Jülich Supercomputing Centre (FZJ, Germany)
LINPACK	Software library for Linear Algebra
NCF	National Computing Facilities (Netherlands)
NCSA	National Centre for Supercomputing Applications (Bulgaria)
NIH	National Institutes of Health, USA
NSF	National Science Foundation, USA
OS	Operating System
PI	Principal Investigator
PRACE	Partnership for Advanced Computing in Europe; Project Acronym
PRACE-1IP	PRACE 1st Implementation Phase project
PRACE-2IP	PRACE 2nd Implementation Phase project
PRACE AISBL	PRACE Association Internationale Sans But Lucrative
R&D	Research and Development
SLA	Service Level Agreement
SME	Small and Medium Enterprise
SSC	Scientific Steering Committee
TCO	Total Cost of Ownership. Includes the costs (personnel, power, cooling, maintenance, ...), in addition to the purchase cost of a system.
Tier-0	Denotes the apex of a conceptual pyramid of HPC systems. In this context the Supercomputing Research Infrastructure hosts the Tier-0 systems; national or topical HPC centres constitute Tier-1
UC-LCA	Universidade Coimbra-Laboratório Computação Avançada (Portugal)

Executive Summary

The objective of this deliverable is to show the results of a pilot impact assessment exercise. This exercise takes as basis the theoretical framework developed in D2.4.1 and D2.4.2, and considers a set of variables that were tagged as short term assessment. In addition, other mid-term variables have been included to provide preliminary data as baseline to propose a way for structuring the data.

The variables considered for the pilot are:

- Success ratio of proposals
- h-index and g-index of applicants
- Resource allocation
- Technical specification of systems available through PRACE and other computer systems made available by PRACE members
- Distribution per job-size and duration
- Training events
- Publications of any type (peer reviewed or not), PhD theses, success stories, etc.
- Project finance structure in terms of additional funding or private/public collaborations
- Software development for scalability development, industrial applications, creation of new collaborations
- Industry participation in PRACE events
- PRACE raising awareness events and media coverage
- Ecological imprint

Given the early stage of implementation of PRACE RI, it is not possible to provide solid indicators on the impact assessed in the previous variables. Nevertheless, the analysis shows baseline numbers to further study a set of selected interesting impact indicators and establish the analysis on the basis of increasing or decreasing trends of quantitative data. For the different variables, an indicative way of analysing future data is provided. Also advises on processes to be integrated in the infrastructure or elements to capture, are provided.

1 Introduction

For the organisation, it is essential to be able to assess results against expectations. In this deliverable, a set of available values from previously identified variables are analysed, providing the baseline for a pilot impact assessment and setting up the basis for future impact assessments for PRACE infrastructure.

Deliverable D2.4.1 [1] described in detail all the aspects regarding monitoring and reporting in PRACE RI. In that deliverable, the management cycle of PRACE AISBL (shown in Figure 1) resulting from the whole monitoring process was described. This cycle is initiated by defining the queries that need to be answered, then identifying the corresponding monitoring variables is necessary, followed by monitoring, reporting and the evaluation process that should lead to adjustments of implementation of PRACE where necessary. Those elements are crucial for asserting the impact of PRACE among users, stakeholders and society in general.
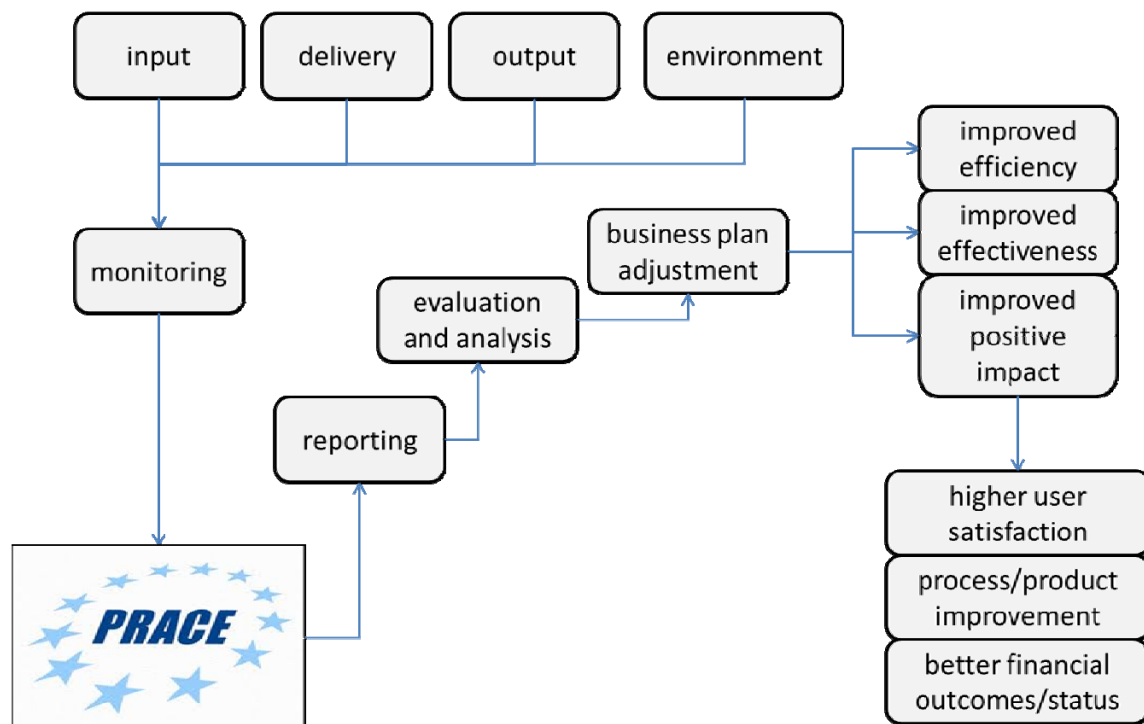
Figure 1: Management cycle of PRACE AISBL

The general set of variables of potential interest to be analysed are classified into different categories.

The first category refers to the focus of the analysis towards the objectives of the infrastructure. In this respect, the assessment can be made to evaluate:

- Efficiency: "Is the organisation being efficient in its operation, achieving as much as it could with its resources?"

- Effectiveness: "Is the organisation being effective in achieving the goals set?"

- Impact: "Is the organisation changing the status quo of the society, economy, science, etc..?"

Some variables can provide useful data for evaluate and improve one or more of the previous elements. This pilot is oriented to provide preliminary data just for the impact assessment

process of PRACE, in order to check regularly what has made a difference to the status quo. In the case of PRACE, it is important to track and verify the impact of the mission defined for PRACE.

The second category refers to the specific stage of the workflow model, and the different values are:

- Input – "what resources PRACE uses in order to create and run its services such as monetary funding and budget, personal resources and equipment."
- Delivery – "the services provided by PRACE, that have a technical character."
- Output – "results obtained with allocated compute resources."
- Environment – "the socio-economic framework in which PRACE operates."
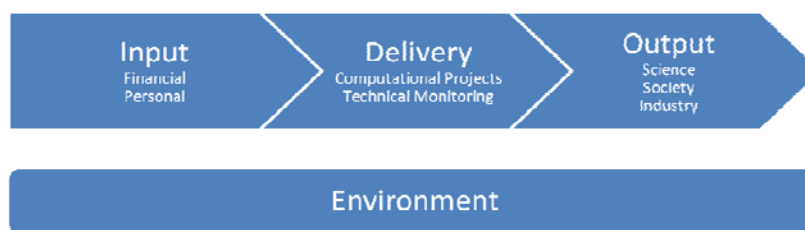
Figure 2: Main components of the workflow of PRACE

In addition, this deliverable builds upon the work done for Deliverable D2.4.2 [2], which develops further a theoretical framework and provides a classification of various impact-related variables. It was clearly identified that impact assessment for PRACE is necessary and possible, requiring a significant amount of operational and organizational effort focusing on a well-established methodology and clear procedures for the analysis of each type of information. In D2.4.2 [2], a subset of the general set of monitoring variables was selected as suitable to assess impact. The different types of impact that variables could assess are specified as:

- Scientific
- Economic
- Social
- Environmental

The fourth dimension by which we classify variables is the nature of the assessment that can be done with their values. In this dimension, the assessment for variables can be:

- Quantitative
- Qualitative

Moreover, according to the timeframe for measuring the impact there is another classification dimension with three main categories considered:

- Short term timeframe
- Medium term timeframe
- Long term timeframe

The current deliverable shows results on the analytical work in a subset of variables (mainly short term timeframe variables of the delivery, output and environment stages referring to scientific, economic and environmental impact) with the currently available data generated by PRACE RI. Variables presented in this report are mainly quantitative and not qualitative. This is due to the fact that within short term timeframe, it is not yet possible to properly identify case studies or success stories. That is why the variables presented are based on defined quantitative targets, or analysis of increasing or decreasing trends, that can be analysed on regular basis.

The recommendations included in D2.4.1 [1] mentioned that for most variables to be monitored by PRACE, validation will be rather light, except for the case of financial variables, where validation is an important step and needs to be performed by external experts. Therefore, those variables are not considered in this deliverable.

The pilot assessment of the short term timeframe variables presented in this analysis is the next step that was advised in Deliverable D2.4.2 [2]. And the set of variables considered for the analysis is:

- Success ratio of proposals
- h-index
- Resource Allocation
- Technical specifications of systems available through PRACE AISBL
- Distribution per job size and duration
- Monitoring variable: Training events
- Publications of any type (peer reviewed or not), PhD theses, success stories
- Project finance structure in terms of additional funding or private/public collaborations
- Software development for scalability development, industrial applications, creation of new collaborations
- Industry participation in PRACE events
- PRACE raising awareness events and media coverage
- Ecological imprint

The following section explains the steps carried out to perform the pilot assessment. Section 3 provides the analysis for each of the analysed variables. Finally, Section 4 summarizes the results and proposes the next steps to build a useful and possible impact assessment mechanism for PRACE RI.

2 Organisation of the pilot assessment

The work has been divided in the following stages:

1- Selection of variables for the pilot

Taking as a basis the results in D2.4.2 [2], we selected a subset of variables. This subset mainly contains variables that were tagged as short term implementation timeframe. For assessing them, it is not necessary to have a wide timeframe perspective. The pilot is mainly focussed in obtaining preliminary quantitative values. In addition, other mid term variables have also been chosen to exercise the capture of baseline data for the future and further the discussion on the type of analysis to be done with the data.

2- Distribution of variables amongst the working group members and discussion on the process for gathering data

For each of the variables selected, the sources of the data in the current status of the infrastructure were identified as well as the process that will be followed to get this data and the necessary permissions or confidentiality considerations.

3- Analysis of the data

Once the data is collected, each variable was analysed following the indications made in D2.4.2 [2]. The analysis made had to be accommodated to the data available at this point in time, and the scope has been limited because of this reason. In many cases, the small amount of data available makes the results not conclusive at this point in time, but interesting enough to keep working on the process in the future.

3 Assessment Results

3.1 Assessment of impact variables

This section describes the assessment process for each of the selected variables.

3.1.1 Success ratio of proposals

For assessing the impact of PRACE based on the success ratio of proposals, we have analysed the information on awarded/requested resources against the evolution of resources available in the infrastructure.

The next graph shows -in blue- the ratio of hours awarded against the hours available. The graph shows that across the first five calls of PRACE, there has been an increasing trend in the success ratio of granted projects. A similar trend is shown for the ratio of projects requested and projects awarded. For the fourth call, the success ratios of both variables are between 50% and 60%. However, in order to analyse properly this scenario, it is important to take into consideration that the amount of resources available has not been constant across the five calls.

In the same graph -in green- the available resources are represented. It can be seen that from the Early Access call to the last call, there is a difference in the amount of resources available of 250%.

This significant difference is relevant for the analysis, and makes evident that there is an increasing trend in success ratio of proposals. This trend can be considered moderate by taking into consideration the large increment in the amount of resources available.
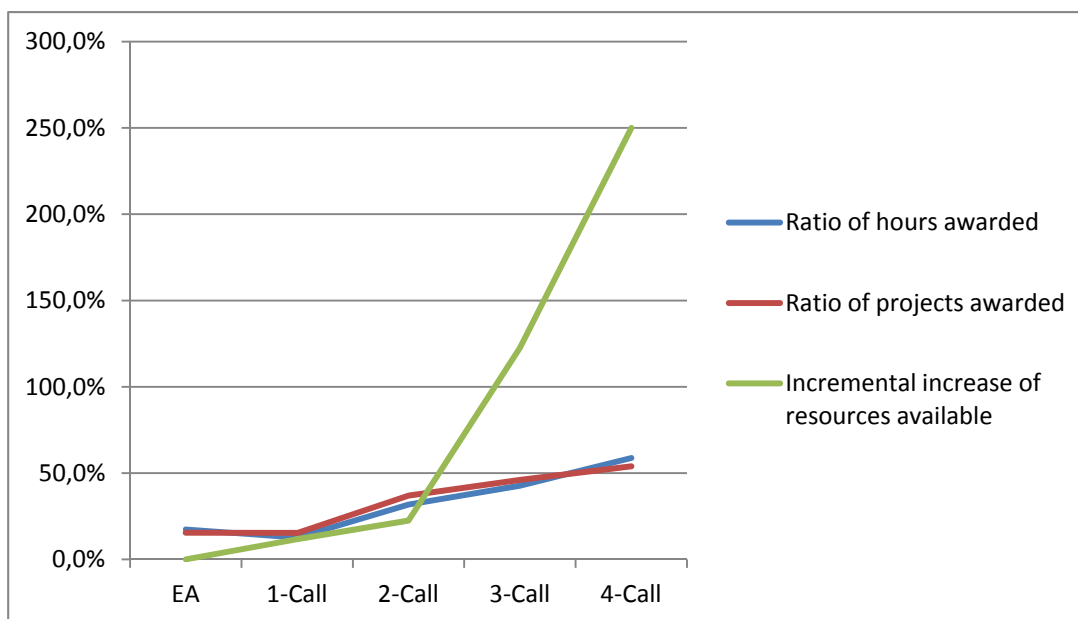
Figure 3: Ratio of increase of projects and resources over time compared with the increase for resources available

Other factors that need to be considered in this analysis are the early stage in which the PRACE RI is in, the fact that the user communities are still adapting to the timings of the calls and to the new resources offered by PRACE.

This analysis will provide straightforward data to analyse results once the infrastructure will be more established (after the initial period) and it will be providing a relatively stable amount of resources. For the period reviewed, the assessment of the reviewed data is that there has been a positive impact in the user communities based on the moderate increment of success ratio compared to the high increment in resources made available.

3.1.2 h-index of applicants

In order to get an overall idea of the scientific merit of the leaders of the projects awarded with CPU hours in the PRACE regular calls, we searched for scientific data in the ISI Web of Knowledge by Thomson Reuters (this is probably the most used scientific database). The data includes:

- Number of scientific papers authored by the PI (indexed in ISI)
- Total number of citations
- Average number of citations per paper
- h-index of the PI

The h-index (as well as the g-index) is directly related to the number of publications and the number of citations to the published work. The h-index (or Hirsch index) is based on the distribution of citations received by a given researcher's publications. A scientist has h-index H if, out of his/her N papers, at least H papers have H or more citations each, and the other (N− H) papers have less than H citations each.

In other words, a scholar with an h-index H has published at least H papers each of which has been cited in other papers at least H times. Thus, the h-index reflects both the number of publications and the number of citations per publication. The h-index works better when comparing scientists, groups or institutions, working in the same field. However, in this study we will consider all scientific areas as a whole. Irrespective of the scientific field, a senior

scientist with a very good publication record and with a very good reputation in the community should have, at least, an h-index equal to 15 and a number of papers above 50. However, it is worth noticing that a very good scientist might have a smaller h-index and a smaller number of papers simply because he/she is young.

So far, there were four regular PRACE calls. The following graph summarizes the averaged data associated with the Principal Investigator (PI) of the projects that were given CPU hours.
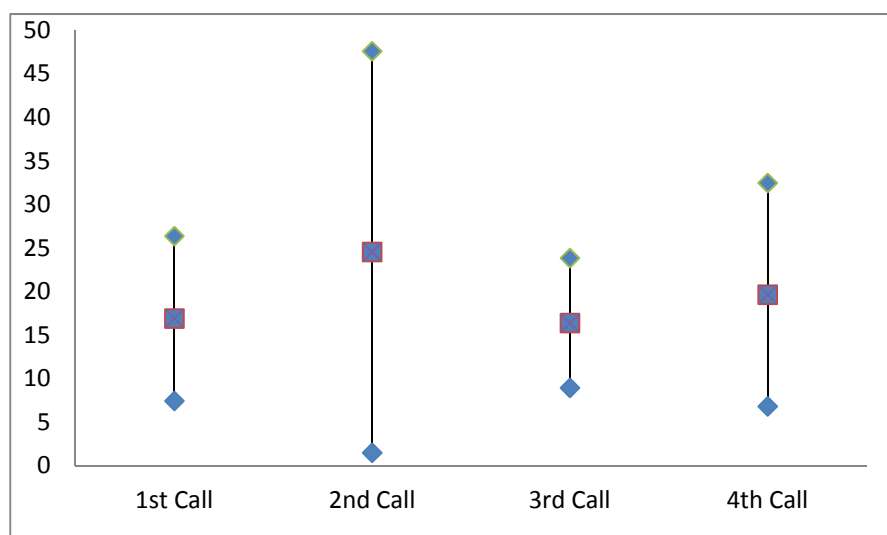
Figure 4: Average h-index +/- standard deviation of the PIs of the projects in the different calls

One should be aware that the information collected in the data base represents a snapshot: the information is dynamic and varies with time.

The above figure shows that the average h-index is always above 15, attesting that PIs are very good scientists. The data associated with calls 1 and 3 is similar and slightly different from the data in the second call which includes a project lead by an outstanding scientist, with more than 500 papers, around 40000 citations and an h-index around 100. The effect of this outlier it is reflected in Figure 4: Average h-index +/- standard deviation of the PIs of the projects in the different calls, where the average is ostensibly increased by the presence of this investigator and the standard deviation indicates the significant difference of values amongst the PIs of the projects in the call.

We recognize that, in all four calls, there is a participation of young scientists, with relatively low h-indexes. Nevertheless, the Access Committee, in view of the quality of the proposals, decided to award CPU time to these projects. This is a point that deserves a special mention in this study: the data reveal that scientific record does not influence the process as much as the scientific merit of the project proposals. We should also stress that the number of scientists with h-index below 10 − young scientists − is relatively small (of the order of 20%) but they are present in the four calls.

The majority of the PIs have an h-index around 20 or more, i.e., they are unquestionably leading scientists in their areas, as must be expected. This is also corroborated by their numbers of papers and by the total number of citations for those papers.

This metric proves to be heavily sensitive to outlier values of the leadership of outstanding scientists in particular projects but in general, within the period analysed, the overall trend of the H-index across the different calls is flat. This indicates that no significant changes are yet registered in terms of the type of users who access the infrastructure.

Three scenarios are possible for the future of PRACE according to this data:

- The trend keeps flat. This data may show that PRACE is not impacting the profile of scientists that make use of PRACE.
- The trend increases. This data may show that PRACE may have impacted in the way scientists make science, and that renowned scientist that might not be used to work with HPC are increasingly embracing HPC as a tool for doing science.
- The trend decreases. This data may show that HPC may become a discipline just used for young or lower profile researchers. This will be a negative result in terms of impact.

In any case, in order to reach any of the previous conclusions, it will be necessary to gather data over several years to make a proper assessment.

3.1.3 Resource allocation

To assess the impact of PRACE based on this metric, it is necessary to have a long term perspective. With the results gathered so far in terms of scientific fields of the projects granted, it is possible to average a basis for future analysis. Results gathered show that except by Mathematics and Computer Sciences, where the applications of HPC are traditionally fewer, the other branches of science have a relatively similar distribution.

The analysis of evolution of future data can reveal two possible trends:

- Similar proportions in scientific usages of HPC. This will reflect a homogeneous development of HPC applications for all fields
- Having a specific area growing significantly compared to others. This may reflect two things:
  - That one area has developed a specific application of HPC with which it is possible to obtain results of significantly higher impact compared to the other proposals from other scientific branches
  - That the usage of HPC is no longer providing relevant results in some scientific fields while in others the activity is kept
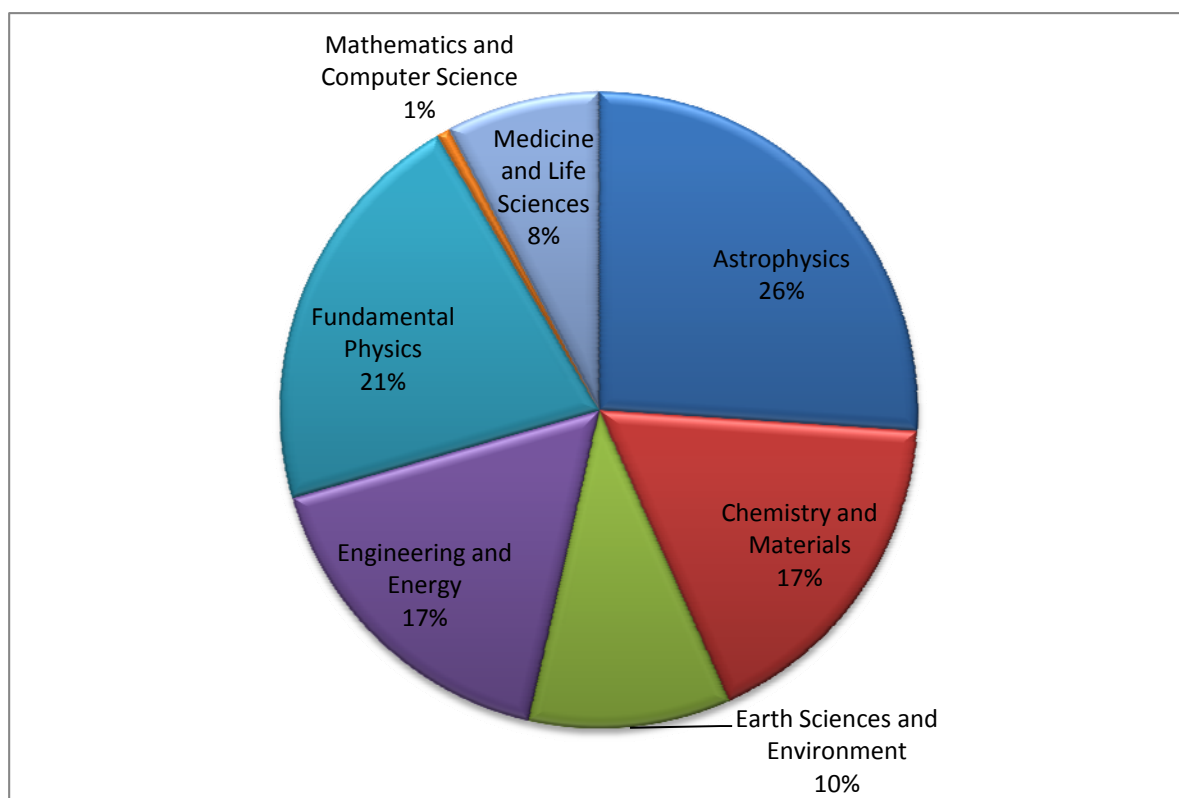
Figure 5: Distribution of resources per scientific areas

PRACE has allocated so far resources to academic scientists. Just very recently, the open research programme for industries was approved. This is the reason why for now the allocation per type of user cannot be assessed. It will be a good practice for the future to see how industrial access increases under the open access or other model.

3.1.4 Technical specifications of systems available through PRACE AISBL and also of other computer systems made available by PRACE Members

The evolution of the performance and capability of PRACE systems is considered a scientific and economic indicator. Given the high cost of having Tier-0 systems and the current economic downturn, if PRACE machines remain competitive compared to those in USA, Japan and China, it will indicate that PRACE has been able to show sufficient relevance to the society to show the economic benefits of investing in this technology. From the scientific perspective, if PRACE systems increase steadily in capability, and continue to be exploited by the scientific community, it will indicate that there is the necessary scientific demand to exploit these resources. This will also show that the scientific community is adapting to the latest facilities and hence this will potentially benefit the community and thus the results it obtains from the most powerful machines.

In order to carry out this analysis, the first element analysed are the position of PRACE systems in the worldwide rankings. The following table shows the ranked position and total Rpeak performance (in Tflops) of the PRACE Tier-0 systems. Note the following data predates PRACE to provide historical perspective. It will be subject to dramatic change with the release of the June 2012 Top500[1] data as this will include HERMIT, SuperMUC and a dramatically upgraded CURIE system. Note the previous upgrade of the JUGENE system between November 2008 and June 2009. It is intended that this table will be extended in time to show the chart the development of the PRACE infrastructure. This will become increasingly valuable in highlighting if PRACE maintains its position as machine replacement cycles take effect.

Machine	Nov-07	Jun-08	Nov-08	Jun-09	Nov-09	Jun-10	Nov-10	Jun-11	Nov-11	Jun-12
JUGENE	2	6	11	3	4	5	9	12	13	25
HERMIT	-	-	-	-	-	-	-	-	12	24
CURIE	-	-	-	-	-	-	86	106	149	9
SuperMUC	-	-	-	-	-	-	-	-	-	4
FERMI	-	-	-	-	-	-	-	-	-	7
Total Tflops	222.8	222.8	222.8	1003	1003	1003	1107	1107	2151	8996

Table 1: Top500 Rankings of Tier-0 Systems

---

[1] http://www.top500.org

Machine	Nov-07	Feb-08	Jun-08	Nov-08	Jun-09	Nov-09	Jun-10	Nov-10	Jun-11	Nov-11
JUGENE	4	4	14	17	18	22	19	29	40	92
HERMIT	-	-	-	-	-	-	-	-	-	61
CURIE	-	-	-	-	-	-	275	225	258	317

Table 2: Green500[2] Rankings of Tier 0 Systems

The Green500 ranking lists the systems in terms of energy usage per floating point operation. It can be considered a supplementary metric to the Top500. It largely groups systems by architecture class but does not consider size per se, indeed large systems can be at an inherent disadvantage.

Complementing this analysis, we consider it interesting to also record the evolution of the number of core hours provided. Once PRACE will have fully deployed its systems, the evolution of the resources made available will be useful for assessing the potential for scientific impact. The analysis of this data will need to take into consideration the evolution of the ratio of successful requests to PRACE.

The following table shows the level resources from each system dedicated to PRACE in the regular calls (in millions of core hours on the given systems). Over time such data will show the growth of PRACE.

	Early Access	1st Call	2nd Call	3rd Call	4th Call	Total
JUGENE	324	362	358	360	299	**1703**
CURIE	-	-	40	161.6	206	**407.6**
HERMIT	-	-	-	207	47.8	**254.8**
SuperMUC	-	-	-	-	200	**200**
FERMI	-	-	-	-	299	**299**
Total per Call	324	362	398	728.6	1051.8	2864.4

Table 3: Number of core hours from each Tier-0 system to regular access calls (in million core hours).

For the sake of future reference, the following table shows the specifications at the time this analysis was made:

Name	JUGENE	Curie	Hermit (Phase 1)	SuperMUC	FERMI
Centre	Gauss Centre for Supercomputing, Jülich, Germany	GENCI, CEA, France	Gauss Centre for Supercomputing, HLRS, Germany	Gauss Centre for Supercomputing, LRZ, Germany	CINECA, Italy
Manuf.	IBM	Bull	Cray	IBM	IBM
Model	Blue Gene/P	S6010 bullx nodes (fat nodes)	XE6	IBM System x iDataPlex (thin nodes)	Blue Gene/Q

---

[2] http://www.green500.org

Name	JUGENE	Curie	Hermit (Phase 1)	SuperMUC	FERMI
		B510 bullx nodes (thin nodes) bullx B505 blades		IBM System x iDataPlex (fat nodes)	
Processors	Power PC 450, 32-bit, 850 MHz, 4-way SMP, L3 Cache: shared, 8 MB	8 core processors, Intel Nehalem-EX X7560 @ 2.26 GHz 8 core, Intel Xeon Intel Westmere	AMD Opteron 6276 (Interlagos) processors with 16 Cores @ 2.3 GHz (with TurboCore up to 3.3 GHz) 32MB L2+L3 Cache, 16MB L3 Cache	8 core processors, Intel Xeon Sandy Bridge-EX 10 core processors, Intel Xeon Westmere-EP	PowerA2, 1.6 GHz
Nodes	73728	360 5040 144	3552	9216 205	10240
Cores	294912	11520 80640 288	113664	147456 8200	163840
Memory (per core) GB	2	4 4 na	2 & 4 (480 nodes)	2 6.4	1
Accelerators					
No.		288			
Model		NVIDIA M2090 T20A			
Inter-connect					
Tech.	Proprietary 425 MB/s in each direction, total of 5.1 GB/s bandwidth per node	InfiniBand	Proprietary	InfiniBand	Proprietary
Details	3D Torus	QDR Full Fat Tree network	Gemini	FDR10 (18 fully non-blocking islands uplinked at a 4:1 ratio) QDR (fully non-blocking)	5D Torus
Linpack	825.5	105	1045	2210	2100

Name	JUGENE	Curie	Hermit (Phase 1)	SuperMUC	FERMI
(Tflops)		1303		65	
		192			
Power					
Max (kW)	2500	2630	2000	3500	1000
Average (kW)	1720	1940		<3000	900
Cooling overhead (not incl.)	20%			10%	15%
Core hours per 4th Call (Million)	300	28	120	200	300
		188			
		na			
Top500 June 2012	25	9	24	4	7

Table 4: Snapshot of technical details of Tier-0 Systems

The data shown can just be considered as a baseline for future analysis when more systems are included into PRACE. The analysis of the rankings can just have value several years beyond the initial period when there will be enough data to show the sustainability of worldwide performance or decrease of competitiveness of the systems over time. It should be noted that a further PRACE Tier-0 system is planned to be deployed by the Barcelona Supercomputing Centre in the near future, however precise details are not available as yet.

The second part of this analysis consists of a log of the Tier-1 systems available in Europe and the evolution over time of the number of such systems, put in context with the evolution of Tier-0 systems. If a greater number of Tier-1 systems emerge, even if the competitiveness of Tier-0 systems decrease, it may indicate that PRACE has impacted effectively the scientific community and promoted the broader adoption of HPC.

The following table lists the Tier-1 systems involved in PRACE since the adoption of the DECI call into PRACE, i.e. calls 7 and 8. Given the large number of systems, extensive technical details are not listed. Over time, this list must expand to accommodate new systems in future DECI calls and also systems that are not made available through DECI calls. Furthermore, future calls will include accelerators as a significant component. It should be noted that the total number of cores hours available at Tier-1 is much smaller (ca. 53million in DECI7 vs. 1 billion for Tier-0 in the 4th call), thus the impact in certain other metrics e.g. ecological impact is minimal[3] relative to Tier-0.

Partner	System	Type	LINPACK [Tflops]	no of cores	DECI7 core-hours	DECI8 core-hours
JUELICH	Juropa	Bull Nehalem Cluster	183	17664	2,900,000	2,900,000
GCS/RZG	Vip	Power6	98		1,150,000	1,150,000
GCS/RZG	Genius	BG/P	48		2,872,000	2,872,000
GCS/LRZ		SuperMUC Migration	60	8200	1,437,000	1,437,000

---

[3] These metrics are not related to the scientific relevance of the results produced using Tier-1 resources, jus to technical parameters. While the scientific impact of Tier-1 usage would be indeed interesting, PRACE is a research infrastructure providing Tier-0 resources and has to focus on their analysis.

Partner	System	Type	LINPACK [Tflops]	no of cores	DECI7 core-hours	DECI8 core-hours
		System				
GCS/HLRS	Laki	NEC Nehalem	112.8		784,000	
		Cluster				
GENCI/IDRIS	BABEL	BG/P	139	40960	6,000,000	6,000,000
GENCI/CINES	JADE	SGI ICE EX8200	267	23040	2,000,000	2,000,000
EPCC	HECToR	Cray XE6	360		7,800,000	7,800,000
BSC	MareNostrum	Cluster	63.8		1,900,000	1,900,000
CSC	Louhi	Cray XT4/5	76.5	10864	2,284,000	2,284,000
ETH Zurich	Monte Rosa	Cray XE6	402	47872		9,000,000
NCF/SARA	Huygens	IBM p575 HydroCluster	50		880,000	
SNIC/KTH	Lindgren	Cray XE6	305	36384	12,749,000	12,749,000
CINECA		IBM SP6	78.7	5376	1,400,000	1,400,000
CINECA		IBM Hybrid Westmere Cluster	120	3288	850,000	
PSNC		SGI UV1000	110	2048	2,747,000	
PSNC - WCNS	Supernova	HP Cluster	30	3840	1,513,728	
SIGMA			80			
NUI Galway	Stokes	SGI ICE EX8200	36.56	380	1,681,920	1,681,920
UYBHM	Karadeniz	Nehalem Cluster		3840		
CASToRC						
NCSA	EA"ECNIS"	BG/P	23.42		2,870,000	
VSB-TUO	Ostrava	86_64 cluster	54			
IPB	Blue Danube	Nehalem Cluster	36			
Total			**2733.78**		**53,818,648**	**53,173,920**

Table 5; Tier-1 systems and contributions to DECI 7 and DECI 8 calls

A continuous monitoring of Tier-1 systems is hence advanced and a time series analysis on the total number of systems, the evolution of their capacity and their territorial evolution is advised in due course.

3.1.5 Distribution per job size and duration

The aim of this indicator is to determine trends in the size and duration of jobs. These trends will show if the usage of systems is adapting to the available resources.

The analysis of this variable permits to assess how parallel the codes are in general. An improvement of this variable over the years is likely to happen as the codes get more and more parallelized but the speed of that improvement will give an indication of the impact of

PRACE on this matter. If a significant fraction of the jobs get close to using the maximum number of cores possible, it would make the case for machines with larger capacity. On the other hand, if almost no jobs get close to the limit, it would indicate either that the codes are not parallelized enough yet to make the best possible use of the machine or that the architecture of the machine is not suited for specific codes or is difficult to program.

It would be good to have this variable at the project level, i.e. tracking for each project awarded access the distribution of jobs according to the number of cores used, but this is unfortunately beyond what is done now in the computing centres. Also, each centre has its way of monitoring this variable and particularly the timespan on which the data is aggregated differs from one centre to the other. Another problem is that in the data provided, there might be no clear distinction between the cycles used by PRACE users and the cycles used by national users (the Tier-0 machines are only partially committed to PRACE).

This is actually the case in the CURIE graph below. On this figure, the distribution of jobs is aggregated over a three months period and for the whole machine (therefore it does not solely represents the PRACE usage).
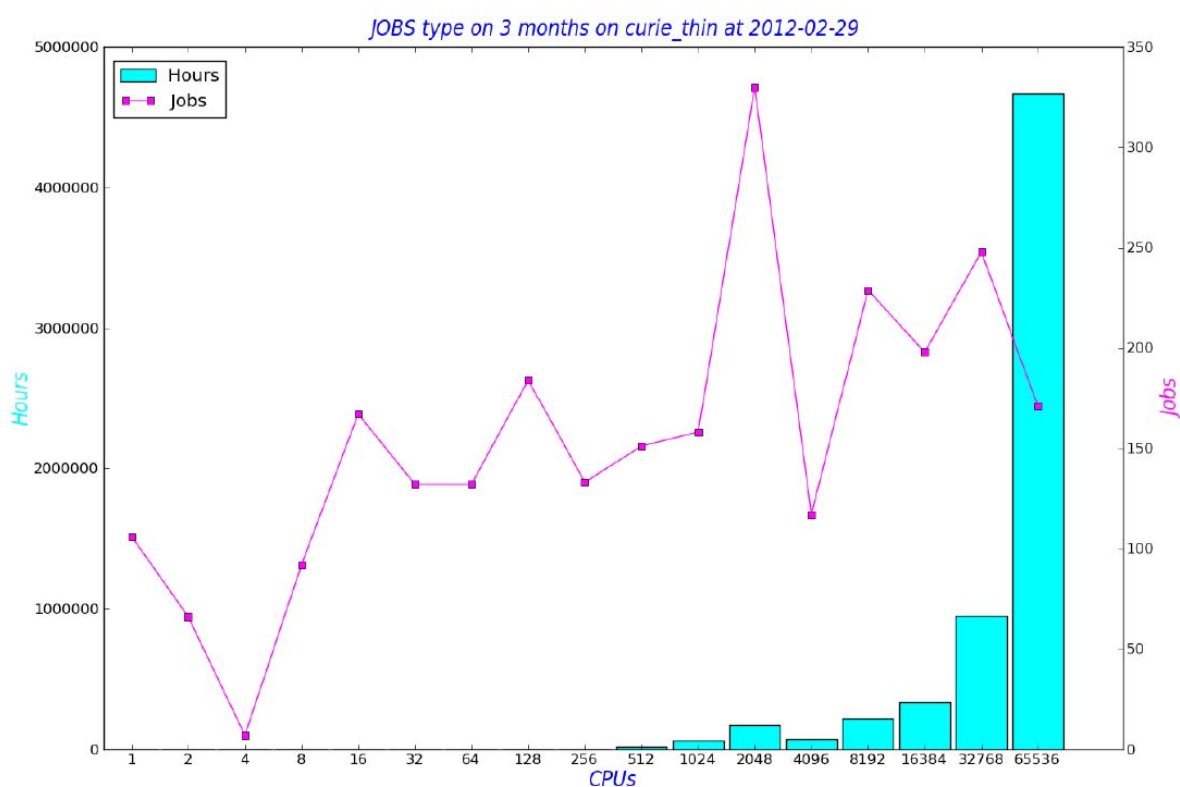
Figure 6: Distribution of jobs and hours by number of cores in CURIE's thin node

The preliminary analysis that can be made on the data shown by Figure 6: Distribution of jobs and hours by number of cores in CURIE's thin node
will be that codes are really exploiting capability computing of the resource. However, it has to be taken into consideration that the data reflects just the results of a "grand challenges" special call to test the machine at its full capacity.

As a conclusion, if this variable is considered by the infrastructure and the computer centres altogether for its integration in the impact assessment process, a clear process should be put in place to be able to regularly get consistent data from the centres and analyse the time series of the average use. The time series will show the potential impact of PRACE in fostering the exploitation of capability computing rather than in capacity computing.

3.1.6 Training events

Training is one of the key activities of PRACE impacting in science. A sustained level of audience or even an increase of it over time may be an indicator of a positive impact of PRACE into the scientific communities. Well performing scalable codes allow treating larger and more complex problems and produce results faster, with better accuracy, which gives a competitive advantage to a researcher. Moreover, training should support the introduction of HPC/eScience into new fields of science and technology. Increased knowledge is finally transferred into economic impact.

Generally, this impact assessment has measured the quantitative impact provided by the PRACE training over the last few years. An homogeneous qualitative analysis mechanism has not been deployed by now. Hence measuring event qualitatively is very difficult based on their different characteristics (seasonal schools and workshops) and rating systems (between PRACE-PP project and PRACE-1IP project). Despite this general difficulty in the PRACE-1IP project there has been progress towards the standardisation of the gathered information about seasonal schools. In that sense, such information on the overall participants' satisfaction from the PRACE seasonal schools is provided here.

The analysis includes the following set of variables:

- **Number and duration of PRACE training events[4]**

From its very beginning, PRACE has identified that a sustained, high-quality training and education programme is a prerequisite to ensuring that the PRACE infrastructure will remain productive. In that sense, it is important to trace PRACE impact in terms of the total number of organized HPC trainings and their geographical distributions throughout Europe.

In that respect, the assessment of the feedback of the training activities can provide important information on how PRACE is impacting the future of HPC to prepare the next generation of European scientists and researchers for utilizing the current massive parallelism of petascale computing, reaching new territories in Europe and training several key as well new HPC community groups both Tier-0 and Tier-1 users.

During the last few years, 18 PRACE training events in 14 European countries have been organized and two events jointly organized jointly with US - NCSA (XSEDE).
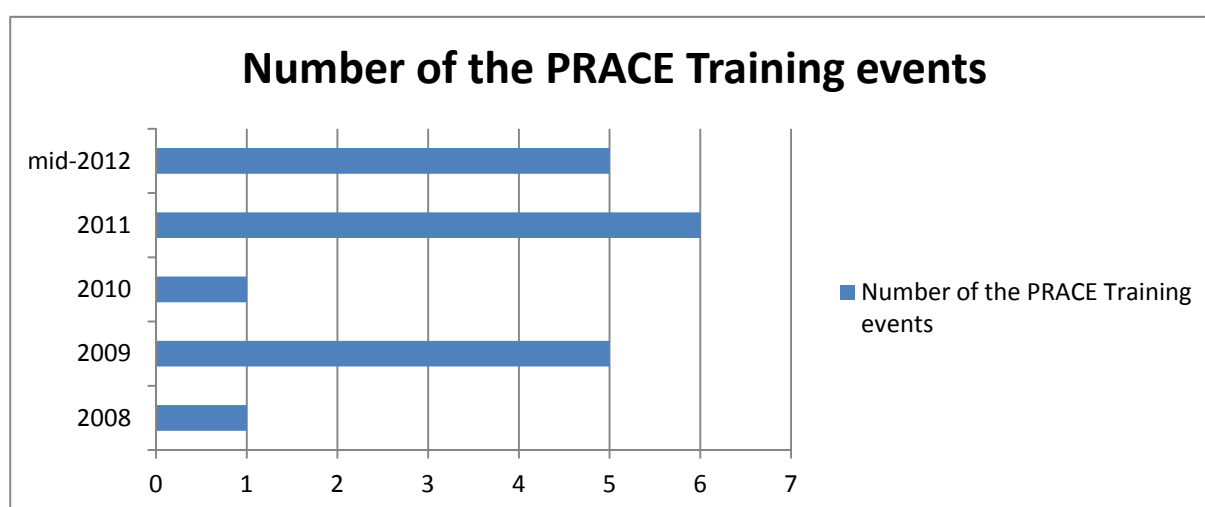

Figure 7: Number of PRACE training events by year

---

[4] For the purposes of this document the variable "PRACE Training events" includes all PRACE Seasonal Schools, PRACE Training Workshops (terminology used in PRACE-PP project) and local trainings supported by PRACE.

Regarding the duration of the courses, the average for the presented trainings is of 3.5 days per event. This duration has proved to be effective and also the preferred one by the training attendants (reflected in the satisfaction survey conducted at end of 2011 [5]).

Starting 2012, the first six PRACE Advanced Training Centres (PATCs) became operational with the mission to serve as European hubs of advanced, world-class training and education for researchers and students in computational sciences, providing and coordinating knowledge transfer activities needed to achieve best utilisation of the PRACE research infrastructure by the community.

In the second quarter of 2012, 18 different PATC events will be organised [6] which means equalling the number of the all other PRACE training events organised up until now in only one year. In other words, only in the second quarter of 2012 through its new training instrument PATCs, PRACE is going to organize the same number of training events than in the whole previous period of its existence.

In addition, the continuation of on-going PRACE Seasonal Schools and PRACE supported local events can significantly boost the total number of reached researchers and students in computational sciences all over Europe.

Although the information shown reflects the activity in the PRACE start-up period, we can foresee a positive impact judging how well received the courses have been until now.

   Number of persons trained in all PRACE training events.

Within the reported period, PRACE organized Seasonal Schools and Workshops reaching 837 people (students, postdoctoral, young researchers, etc.) in different European countries, reaching an average of 47 participants per event. The aggregated statistical information does not include local training events with the direct support of the PRACE projects as such statistics will be available only later this year.

Figure 8: Number of PRACE trained people per year

This result can be seen as a building block in PRACE efforts to prepare the next generation of European scientists and researchers for utilizing the current massive parallelism of petascale computing architectures containing hundreds of thousands of cores and providing them with some insights about the near future of emerging technology, programming languages and models.

In the future, it will be interesting to investigate the correlation between the total number of trained people by PRACE and an increasing number of excellent project proposals for Tier-0

and Tier-1 calls leading to higher competitiveness and scientific excellence of the supported PRACE projects.

In order to do this, the information that is collected during the events should be extended including the way feedback from participants is measured.

Number of persons trained in PRACE Seasonal Schools

PRACE Seasonal Schools are the main training activities directed at improving "application development" and "advanced" competencies. **Ten Seasonal Schools** geographically distributed have been held during the course of the reviewed period providing **an average number of 47.5 participants per Seasonal school or 475 people in total** with an opportunity to enhance competencies that researchers require to exploit the PRACE Tier-0 ecosystem. Starting with one Seasonal School in the period between 2008 and 2010, there is more than a doubling of the PRACE efforts (four Seasonal Schools) only in 2011 and three more in 2012 (until June). Overall, the PRACE Seasonal Schools can be considered as being completed very successfully.

Figure 9: Number of trained people in PRACE Seasonal Schools per year

In PRACE-1IP project in order to demonstrate the qualitative impact provided by PRACE seasonal schools, feedback from the participants has been gathered. The same feedback form has been used for the different events.

A conclusion from the data is that we can say that PRACE is having a positive impact in scientific communities judging from the increasing trend in the demand of the increasing amount of training events prepared.
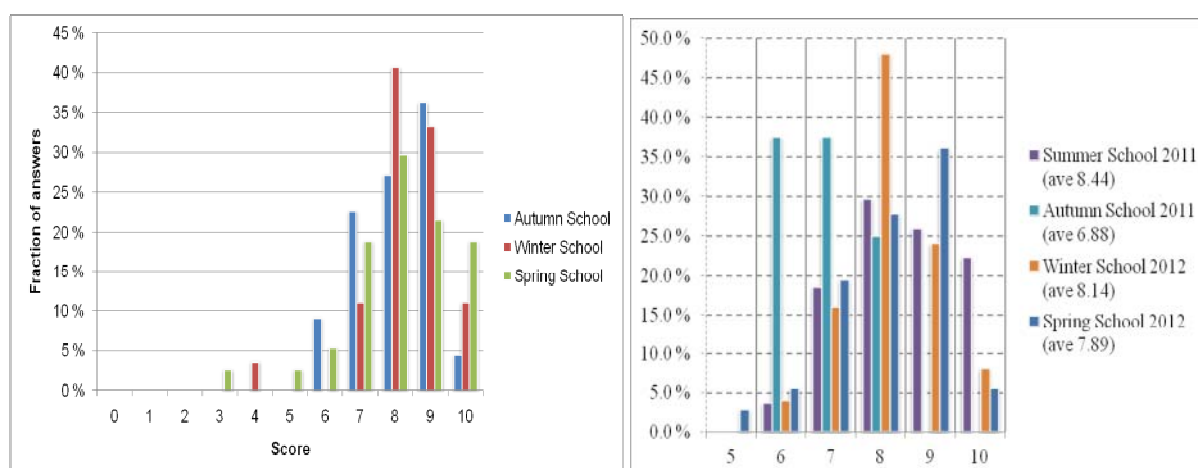
Figure 10: Overall impression scores for seasonal schools in PRACE-1IP [5]

The high quality of practical arrangements and programme content is visible in the feedback. The average scores in response to the question *"Overall, how would you rate this school? [0 = waste of time, 10 = excellent]"* gives an indication on the effectiveness of the operation of the school, but also on the impact in the audience, that rated the contents in a very positive manner. All eight seasonal schools received similar overall rating above 8 from the participants, with only the Autumn School 2011 receiving a slightly lower score.

Additionally, in the future PRACE should not limit itself in measuring only face-to-face training events and needs to pay higher attention to:

- User activity in Training Portal – Number of visits/hits, feedback
- Training material made available through the PRACE Training portal: amount of material by scientific domains and user ranking of material

There should be a focus on PRACE capability to reach out to a vast number of European HPC users (far beyond the face-to-face meetings) via the pervasive power of Internet. Furthermore, there is a necessity of measuring the PRACE ability to contribute to HPC users skills and knowledge obtained through its Training Portal including the provision of e-learning courses. Such information can be collected through regular monitoring of the PRACE training portal and/or questionnaires and other forms of feedbacks.

In PRACE-2IP project and PRACE-3IP project, it will be also convenient to extend training analysis including results from a normalized participant satisfaction form for all PRACE training events to assess not only the amount of participants (that is heavily related to the number of events organized) but also on the perceived utility.

3.1.7 Publications of any type (peer reviewed or not), PhD theses, success stories

In order to assess how PRACE is impacting science, a straightforward indicator is the monitoring of the scientific production over time. With this data it will be possible to extract conclusions depending on the trends that the data reveals.

For this pilot assessment, the follow up reports of the PRACE users have been used as input. Although the data is incomplete, it can give a quantitative idea on the magnitude of the scientific reporting of results based on using PRACE resources.

---

[5] Based on the information provided in the Deliverable 3.2.2 and Deliverable 3.2.5 of the PRACE-1IP projects

The following table summarises the data gathered up until now, that includes reports of projects granted with the Early Access Call, First and Second Access Call closing in January 2011 (for the later calls reports will be requested at a later stage).

	2010	**2011**	**2012**
Invited Talks	11	68	26
Publications	2	31	20
Reports	0	0	29
PhD Thesis	1	3	6
MSc Thesis	0	2	1

Table 6: Tier-1 systems and contributions for DECI 7 & 8 calls

The previous table shows a baseline information that will need to be completed in the future when more reports will be gathered. It is foreseen that production in 2011 and 2012 will increase significantly with the resources granted with the allocations from the three calls that are being used at the time this deliverable is being written.

The analysis of further data collected will reveal the scientific impact depending on the trend of the scientific production. If this increases over time, data will be revealing a positive impact, and if it decreases, it will show negative impact. Nevertheless, for an accurate assessment, a qualitative measurement will also be necessary. It would be recommended to ask for the impact factor of the publications and the audience of the talks in order to be able to evaluate with more precision if production increases not only in quantity but also in quality.

A follow/up of old projects is also advised in order to assess the continuity of the projects, which accessed to PRACE resources.

3.1.8 Project finance structure in terms of additional funding or private/public collaborations

This variable provides information on the engaged different additional funding sources or public/private collaborations during the life cycle of a given project supported by PRACE.

The following data extracted from the application forms of the projects shows the ratio of projects (until the 3rd regular call) that have received some type of supplementary funding:

Period	**Projects with additional funding (%)**	**Projects with additional FP7, ERC or ESFRI funding (%)**
Early Access Call + (1st,2nd,3rd)Regular Calls	71,7%	39,5%

Table 7: Ratio of awarded projects from the different PRACE calls that have reported additional funding

Though this data corresponds to the start-up of PRACE, it can confirm that PRACE is complementing national and European funding and is contributing to the reinforcement of the development of science in Europe. Nevertheless the percentages given above should be seen

as an indication and a more thorough analysis needs to be done when more data is available to see the trend in these percentages.

In general, an increase of co-funding can be interpreted as a positive impact since it will indicate that an increasing amount of projects that involve HPC are being evaluated positively by other different competitive processes of funding.

3.1.9 Software development for scalability development, industrial applications, creation of new collaborations

This variable focuses on the description of the existing software and its development for usage by the European academia and industry. In this respect, we have analysed the data on the work of PRACE in terms of code development. This is shown in the figure below.
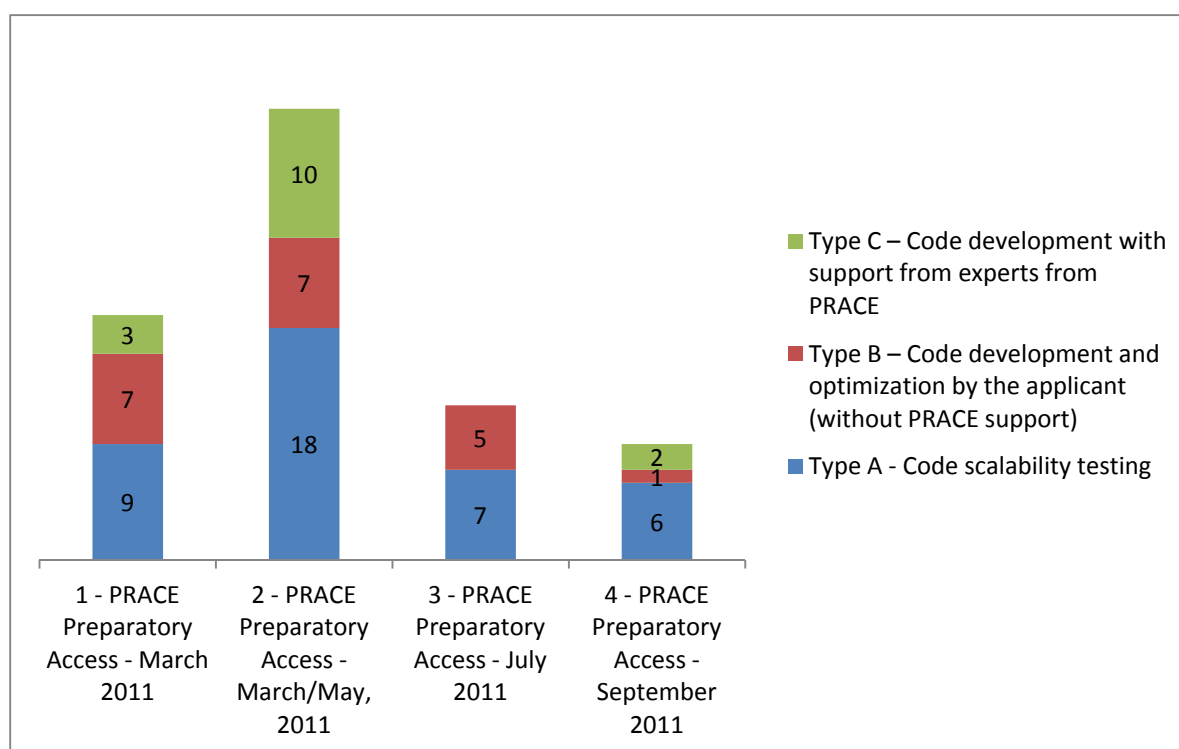
Figure 11: Amount of code developments with support from PRACE

The data above, shown against a timeline, does not show any relevant quantitative trend. In order to assess the exact impact of such activities, there is a need to collect information on the application of these and other software codes (i.e., one needs to look at e.g. the scale of each project and where its results will be used further). This would mainly be anecdotal evidence, or case stories and in some cases the codes do not present enough maturity for their impact to be assessed.

It is recommended that PRACE establishes a resource in order to gather such success stories and assess their impact on completion and the implementation of results. Such information will be qualitative rather than quantitative in nature. This analysis could take the following form, for example:

- A description of how a solution enabled due to PRACE has resolved (or is resolving) significant (to date unresolvable) issues
- Showing the progress of important technological advances made possible by the existence of PRACE

- Linking a number of applications, industrial developments or collaborations to related activities within PRACE.

The above could be made available on the PRACE website and other PRACE publications. This variable, however, assumes the establishment of a long term programme within PRACE RI for software support. If this programme is not established, the impact of PRACE according to this variable will be very difficult to measure.

3.1.10 Industry participation in PRACE events

PRACE Industrial Seminars

The following chart quantifies industry's participation in the three PRACE Industrial Seminars to date:
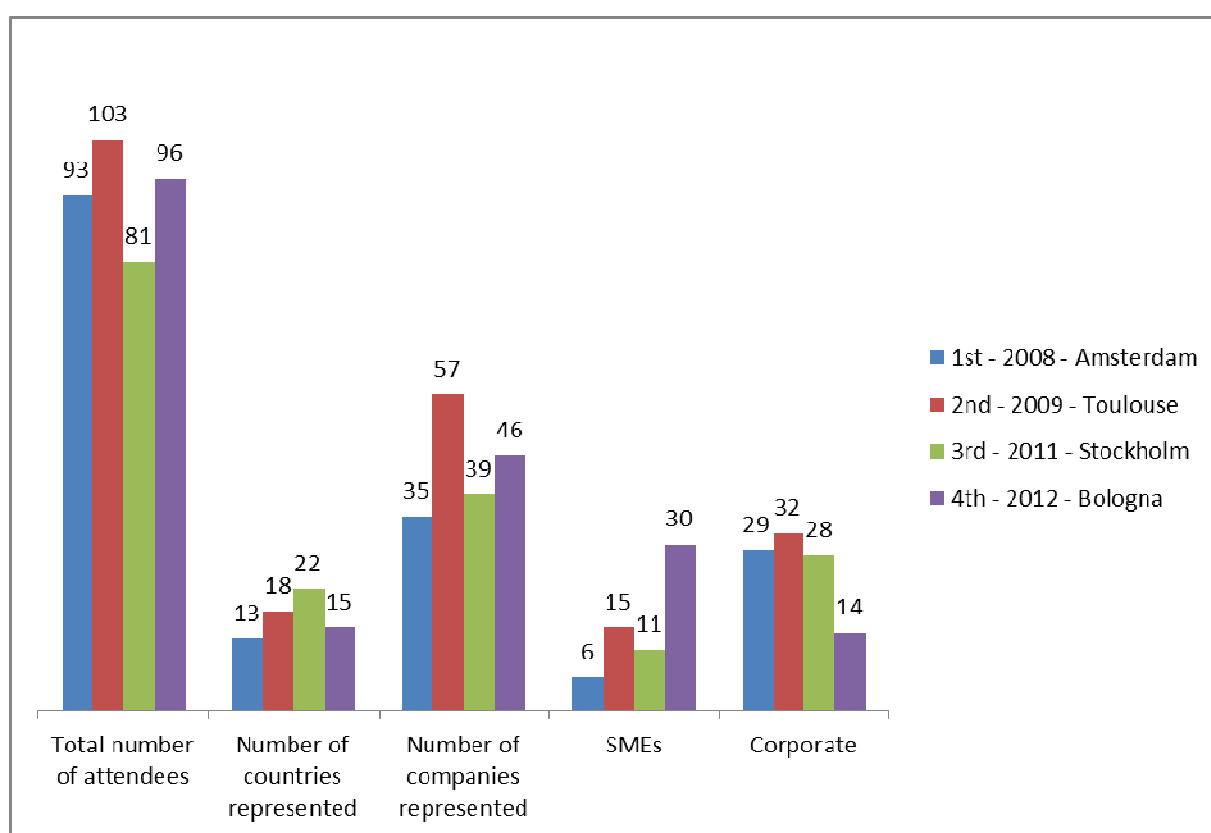
Figure 12: Attendees in the different PRACE Industrial Seminars

The graph clearly shows that the PRACE Industrial Seminars continue to attract a solid number of industrial participants, in terms of the total number of attendees, the countries and the companies represented. There is a noticeable growing trend of the number of SMEs attending the event, which is due to the work of WP5 in PRACE-1IP with that type of companies.

Also, at each of the four seminars, at least 60% of attendees represented the companies that attended a PRACE Industrial seminar for the first time (e.g. at the 4th Seminar, 73% of all attendees).

The 4th PRACE Industrial Seminar held in Bologna 16-17 April2012, was a culmination of the Work Package's efforts in creating a fruitful relationship with industry. The following was achieved:

- The majority of companies were engineering companies working in the simulation area:
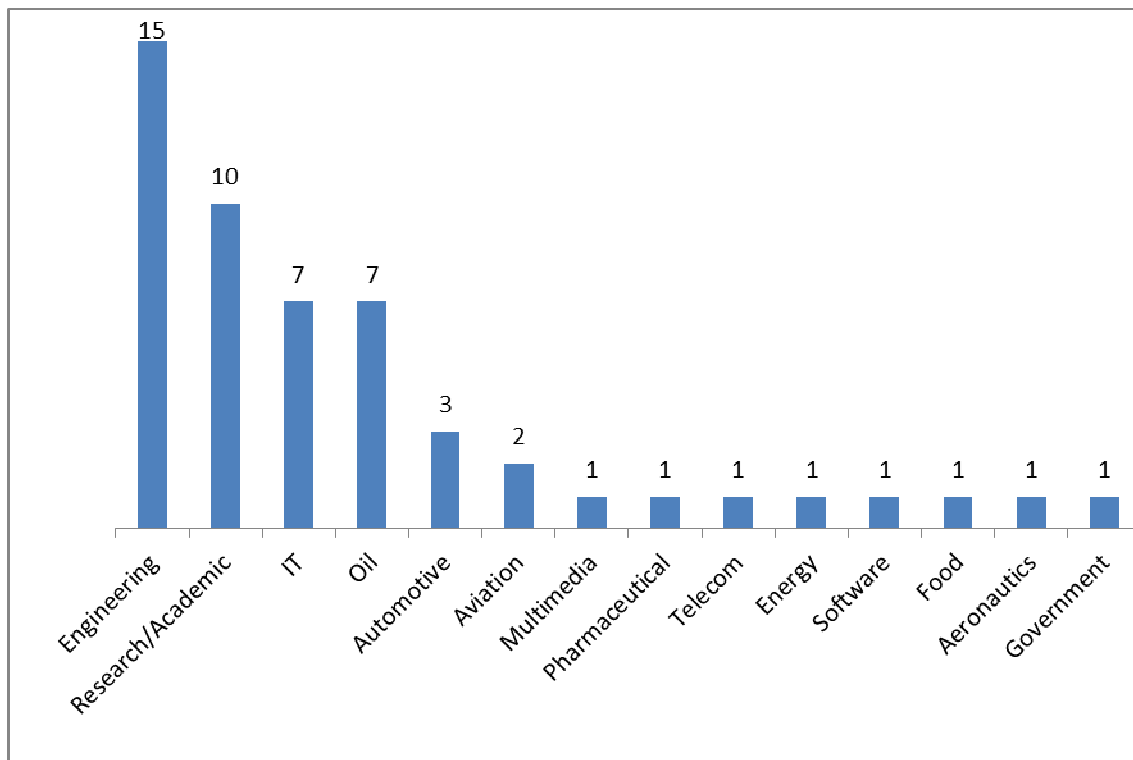
Figure 13: Distribution of attendees per industrial sectors

- Also, the seminar has been rated positively: 71.4% rated it as productive, and 28.6 as very productive. No participant found it non-productive.
- For the first time, PRACE awarded the Most Innovative Industrial Application in Europe. This competition attracted five companies.

HPC Vendor participation in PRACE Infrastructure Workshops

1. First Workshop:

   The Swiss National Supercomputing Centre – CSCS in collaboration with CEA of France and BAdWLRZ of Germany organized the first European workshop on HPC infrastructures on 2–4 September 2009, in Origlio, Switzerland. This event brought together for the first time experts in construction and operation of supercomputing facilities from Europe and around the world, including members of the PRACE project.

2. Second Workshop:

   CEA, in collaboration with CSCS (the Swiss National Supercomputing Centre) and BAdW-LRZ (Germany) organized this second European Workshop on HPC Centre Infrastructures on 6-8 October 2010, in Dourdan, France – near Paris.

   The organisation effort was partly supported by PRACE Preparation Phase Project (Extension Phase Work Packages 7 and 8, from January to June 2010) and partly by PRACE-1IP (Work Packages 8 and 9, from July to October 2010).

3. Third Workshop:

BAdW-LRZ (Germany) organized the Third European Workshop on HPC Centre Infrastructures on September 21-23, 2011, in Garching, near Munich, Germany (Deliverable 8.3/PRACE-1IP).
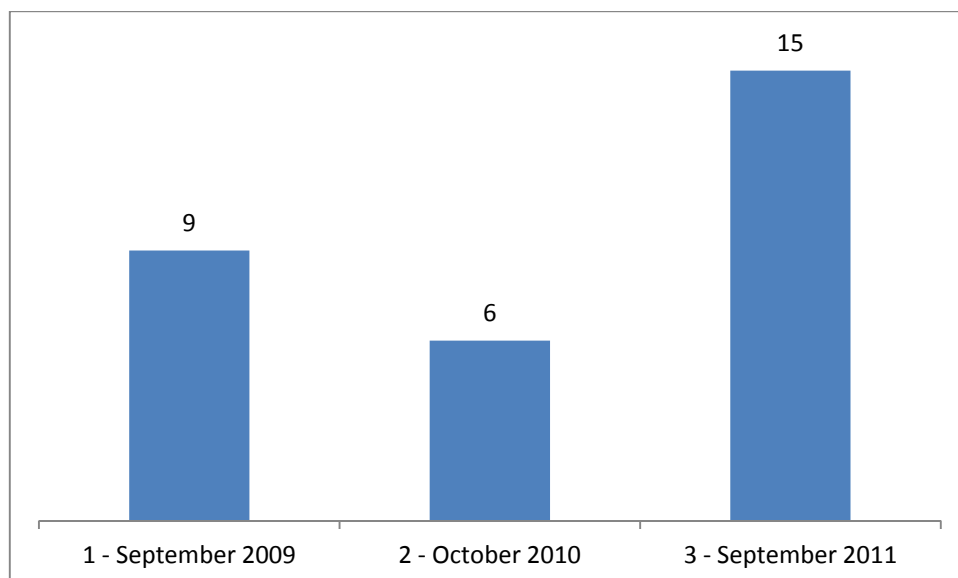
Figure 14: Number of industrial vendors participating in PRACE infrastructure workshops

The graph above shows an upward trend reflecting the growing interest of HPC vendors in collaboration with PRACE. If this trend is sustained over time, that may be an evidence that PRACE will be impacting the industry.

3.1.11 PRACE raising awareness events and media coverage

The visibility of PRACE to both scientific communities and general public is a step leading to the creation of an informed overall perception and an increase and coherence understanding for the role and following impact of HPC on science, society and economy.

In this sense, the main objective of this variable is to evaluate the implemented PRACE platform for dissemination and explanation of important HPC achievements including results obtained by PRACE allocated resources based on a set of channels utilized for reaching a wide array of interested parties.

Even though this variable can encompass a wide scope of events (industrial events, seminars, training events and many others), in order not to replicate the data that is already present in other variables analysed, we have just focused here in the analysis of the impact of the PRACE web site up until today, and also on the PRACE visibility on the two main international exhibitions in the area of HPC.

Figure 15: Number of contacts gathered at International Exhibitions by PRACE

During the last two years, PRACE has paved the way to position and brand itself at the international stage as the leading European HPC infrastructure. Whithin this period, PRACE has taken part in the well-know international exhibitions in the field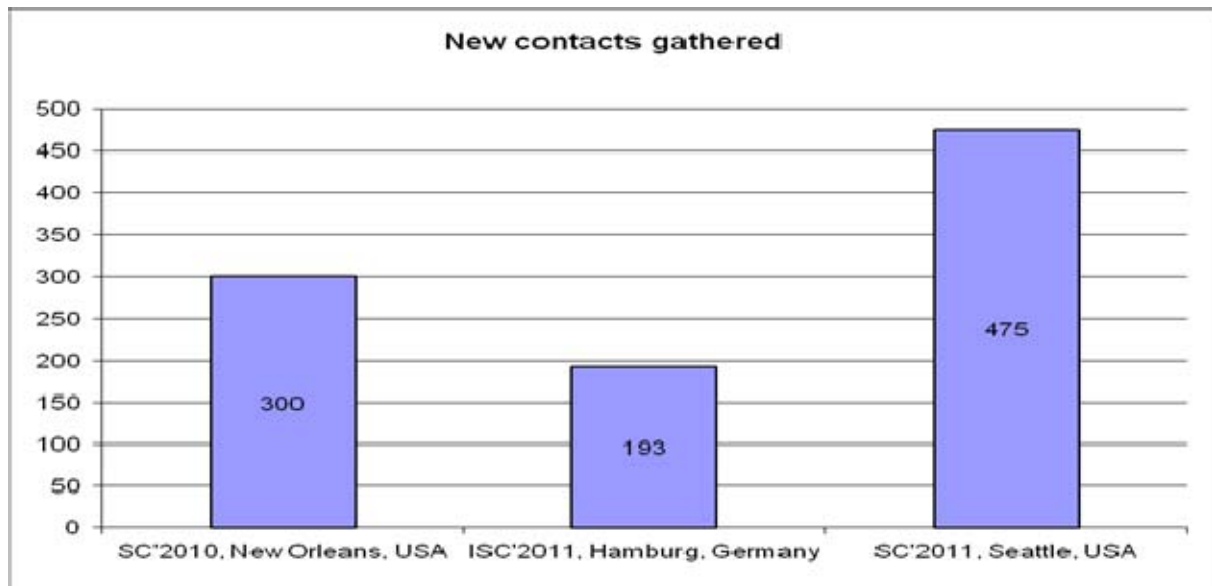 of HPC such as the annual exhibitions - International Supercomputing Conferences (usually held in, Germany) and Supercomputing Conferences in US. Only with its participation in three such exhibitions, PRACE succeeds to reach out to almost 1000 new contacts ranging from:

- Worldwide HPC communities' representatives working in the field of enabling high performance computing applications, deploying and operating of computing and data management infrastructures and advancing scientific discovery:

- Industrial and Vendor representatives working on the design and production of high performance computing systems and/or applications enabling and development as well as fostering the technology transfer – turning the advanced scientific discovery into new products and services:

- European authorities at EU or national level including policy makers, decision bodies and/or funding agencies responsible for HPC policy development and operational implementation:

The accounting of the contacts have been made through the analysis of the data collected during the exhibitions, where PRACE event team uses the retrieval machine to scan the badges of the visitors at the booth and the contact information is kept in the PRACE CRM.

The overall impact of the PRACE participations on these international events should be measured by the strengthened reputation among the main stakeholders and the established new channels for mutual beneficial partnerships on concrete scientific and engineering problems. Data in future events should be also recorded and compared to previous one. A regular decrease in the number of interested participants may show a decrease in the impact, while a steady value or an increment may show a positive impact to be analysed in the context where the data is captured. There is no doubt that the best way to reach out to a global audience in today's information age is through Internet as a global media. The number of unique visits on PRACE AISBL web site per day, as it can be seen in the graph below (showing data since February 2011), has been gradually increasing, and nowadays the PRACE website has an average of 600 unique visits per day.

Figure 16: Number of visitors to the PRACE website since January 2011 until March 2012

The increasing trend is a clear indicator of the impact that PRACE is achieving in terms of visibility and the way PRACE would like to communicate with its users and general public. The website is an integrated part of the PRACE efforts to provide its users opportunities to apply online for computational resources and find practical information about the current PRACE scientific results and training opportunities.

The data reveals that punctual facts, as the announcements of the PRACE Tier-0 calls, have a significant effect in the access data. The steady increase of the visits to the PRACE web site will be a clear indicator on the widening of the audience of PRACE. This audience will mainly be a reflection of the increase of PRACE user base, however it may be also reflect the increasing awareness by other types of visitors that may want to get to know more about the HPC infrastructure that is impacting economy, science or society.

In order to know more on the type of profiles that access the PRACE website, we have analysed the access by website sections. The data up until now reveals the following distribution:

1. HPC access – 34.4%

2. Media – 18.9%

3. PRACE RI Partners – 15.2%

This data characterizes the typical user as a scientist interested in getting access to the infrastructure, but there is also a significant amount of visitors interested in general news or in knowing about the partners of PRACE. The evolution of this data over time may reflect an increase of this type of general visitors (not interested in knowing about HPC access but on reading news about the advances of the infrastructure. This may be an indicator of a societal impact if this change of typical user is accompanied by a general increase in the PRACE web site access.

3.1.12 Ecological imprint

The ecological imprint of HPC is directly related with its energy consumption and more precisely with the performance that systems can get with the energy they need. The following analysis takes into account the technical data available of the systems that PRACE is running. The following table shows initial snapshots of power usage figures based on available figures:

Power related data					
Name	**JUGENE**	**Curie**	**Hermit (Phase 1)**	**SuperMUC**	**FERMI**
Max (kW)	2500	2630	2000	3500	1000
Average (kW)	1720	1940	na	<3000	900
Cooling overhead (not incl.)	20%	na	na	10%	15% (expected) 30% (worst case)

Table 8: Gross power usage of Tier-0 Systems

Additional public statements related to environmental impact of PRACE Tier-0 systems:

JUGENE

- JUGENE has a maximum power consumption of 35 kW per rack giving a total of 2.2 - 2.5 MW plus a further 20% for cooling. However the average load when executing applications is 23.9 KW. The idle power usage is 9.9 KW per rack.
- The system is air cooled at a rate of 60.000 m³/h and a temperature of 16°C. This requires 69 fans per rack or 4968 in total.
- The air in turn is water cooled with a maximum inlet temperature of 17°C, 24-27°C at the outlet. Warm water is used to increase the inlet temperature from 6°C a cold water supply.

SuperMUC

- SuperMUC will use a new revolutionary form of warm water cooling developed by IBM. Active components like processors and memory are directly cooled with water that can have a temperature of up to 45 degrees Celsius. The "High Temperature Liquid Cooling" together with very innovative system software promises to cut the energy consumption of the system. In addition, all LRZ buildings will be heated re-using this energy.
- SuperMUC will be housed in the recently expanded Compute Cuboid of LRZ with a target PUE of 1.1

FERMI

- Fermi will reuse existing water cooling infrastructure and free cooling.
- Power distribution infrastructure (e.g. UPS and generators) will also be reused.
- The PUE value will be 1.3 (worst case) but in practice should be roughly 1.15.

HERMIT & CURIE

- No further environmental data available.

Ecological impact relating to electrical power

The principle recurrent ecological impact of HPC systems is their intensive use of electricity for both calculation and the resulting need for cooling. The 'headline' impact of electricity generation is the associated production of CO_2 by the generation technologies which burn fossil fuels. This section aims to give a brief comment on the impact of the electrical power generation in Europe and to relate this to 'end user' computation. Each country has a mix of power generation methods and each system have a given performance per watt. The following table provides a guide to the CO_2 intensity of the existing PRACE infrastructure.

Machine	Average power (MW)	Cooling overhead %	Total power (MW)	Pflops	MWh per Country[6]	Tons CO2 per Country[7]	Ton C02 /MWh	Kg /MWs	kg/MWs per Machine	Kg /Pflops
Jugene	1.72	20.00	2.06	1.00	6.4E+08	3.9E+08	0.61	0.17	0.35	0.35
Hermit	1.75	20.00	2.10	1.08	6.4E+08	3.94E+08	0.61	0.17	0.36	0.33
SuperM UC	3.00	10.00	3.30	2.28	6.4E+08	3.94E+08	0.61	0.17	0.56	0.25
FERMI	0.90	15.00	1.04	2.10	3.6E+08	1.5E+08	0.43	0.12	0.124	0.06
CURIE[8]	1.94	20.00	2.33	1.60	5.5E+08	4.8E+07	0.09	0.02	0.06	0.04

Table 9: Tier-0 derived CO2 levels

Thus if for sake of argument, we take the current PRACE capacity as being the total capacity of the above systems the total performance is 8.05 Pflops. Allowing for the differing power generation systems of the nations involved the current CO_2 production is 180g/Pflops.

It is recommended to review this data in the future after deployment of newer Tier-0 systems and upgrades to the existing ones or hosting facilities to analyse how the CO_2 production evolves. A decreasing rate of CO_2 per Pflop ratio will indicate a decreasing ecological negative impact, although the total number of CO_2 generated every year will need to be also considered in the analysis since the number of compute hours provided by PRACE will likely increase over time.

---

[6] Source for power generation and CO2 production: http://carma.org
[7] Note: while individual centres may be provisioned by power from a given source of a given type with a specific C02 intensity only the national figure is used here, thus accounting for base load issues.
[8] Estimated value for cooling in case of CURIE, 20%

4 Conclusions and next steps

4.1 General Impact Assessment based on preliminary results

The pilot impact assessment performed for PRACE is an exercise with the objective of illustrating the type of analysis that could be performed and the type of conclusions we can extract from it. PRACE aims at having a big general impact and in order to assess the success of this mission it will be necessary to gather and analyse further available information and evaluate it according to a set of criteria and thresholds. According to a Balanced Scorecard type assessment, it will give the possibility to feed the management of the organisation with valuable data in an operative format for redefining the PRACE strategy and the necessary actions to carry it out. However, given the early stage of this pilot, the results obtained are only indicative as to express a verdict using a Balanced Scorecard as advised in the methodology described in D2.4.2 [2].

Although the majority of results obtained in the current pilot are not conclusive enough to provide a complete assessment, they can give us some hints on the type of thresholds to set up for further analysis. In future assessments, a Balanced Scorecard should be offered as an executive conclusion of the assessment, but because of the named limitations of the current research given the early stage of the infrastructure, we present the conclusions as a summary of analysis and recommendations made for each one of the variables:

Success ratio of proposals

The success ratio of proposals for PRACE is currently slightly above 50%, which is a higher ratio than what would be desirable. However, this result, as well as the subtle increasing trend on the ratio across the different calls, is explained by two facts: the start-up age of PRACE and the exponential increment of resources made available across the calls. It is expected that while PRACE calls become widely known and are scheduled well ahead in the agenda of researchers, and the type of resources available become well known, the level of oversubscription is expected to increase.

H-index

The average h-index of the PI of the projects granted in PRACE (around 15) characterizes the user profile of PRACE resources as a young researcher. This type of profile probably reflects that the HPC discipline is rather young, and senior researchers might have not been sufficiently exposed to it as to have enough knowledge to lead a project based on it. This average profile has shown to be stable across the different calls. Future changes in this trend may provide indicators of status-quo changes from the scientific perspective. If the trends increases, it may reflect a positive impact by having an increasing amount of top researchers joining HPC related projects, or a fast growing profile of the HPC-user researchers. And if the trend decreases, it may reflect a negative impact, by showing that top researchers do not participate in this type of projects and that researchers do not repeat requesting resources to PRACE over time.

Resource Allocation

The data obtained sets a baseline of distribution of resources across different scientific disciplines. Future data on next calls will provide information to assess the impact of HPC (and indirectly of PRACE) on specific areas by notable and sustained increase or decrease of projects granted in these areas.

Technical specifications of systems available through PRACE AISBL

The data obtained sets a baseline of technical resources rated against a global positioning. With future systems made available to or by PRACE, it will be possible to see in a timeline fashion if PRACE maintains, reduces, or improves its global positioning. Combined with other impact assessment indicators, this result will be valuable to assess if the strategy to maintain, increase or decrease technical leadership worldwide, is aligned with the impact obtained. The analysis also includes the monitoring of Tier-1 systems as a complimentary factor for the previous analysis, since even if PRACE may show no possibilities of maintaining a high degree of worldwide technical capability competitiveness, its impact may be reflected in an increasing number of Tier-1 systems spread across Europe.

Distribution per job size and duration

The analysis proposed based on counting the number of processors that use each job in PRACE projects would provide a view on how the capability requirements for Tier-0 HPC projects increase over time. A significant increase at some point in time may reveal a scientific impact of PRACE by showing success in the promotion of high-scalability of codes. However, the specific analysis will have to be adapted to the context where the data is framed where many other explanations may be possible.

The working group considers this variable of high interest, however the decision to include it in the monitoring and reporting procedure must be the result of a discussion between PRACE and the computing centres. In case of acceptance, since this data is not being currently collected by centres in a homogeneous way, it will also be necessary to agree on a specific implementation of the procedure.

Training events

The data of the training events organized by the different PRACE projects have been analysed showing that PRACE is having a positive impact in scientific communities judging from the increasing trend in the demand of the increasing amount of trainings prepared. Additionally, in the future PRACE should not limit itself in measuring only face-to-face training events and needs to pay higher attention on:

- User activity in Training Portal – Number of visits/hits, feedback
- Training material made available through the PRACE Training portal: amount of material by scientific domains and user ranking of material
- User satisfaction captured in normalized forms

Publications of any type (peer reviewed or not), PhD theses, success stories

The regular follow up process set by PRACE to its users has gathered data on the invited talks, publications, reports, PhD theses, and MSc theses produced thanks to the results obtained with the usage of PRACE. The data collected up until today includes results of the resources used until the second regular Call (closed in January 2011). The total number of items have been summed for each category and presented for each year. The current amount of data is not conclusive since it just represents the start of the collection of the first scientific results. But a periodic analysis like the one started will show how PRACE is impacting in the scientific production of HPC in the user community.

Project finance structure in terms of additional funding or private/public collaborations

The data provided by PRACE users in their application forms indicates that there is a significant number of projects that count with additional funding (71.7%. 39.5% of them have European funds support). Future data collected will show if this support is increased over time or decreased. Depending on the specific context, this data may be an indicator of a positive

impact in case support is increased, indicating that HPC related projects are acknowledged by an increasing amount of funding authorities as valuable enough as to support them.

Software development for scalability development

The data analysed on the codes development with PRACE participation does not show valid information to assess the impact based on quantitative criteria. It is recommended that for the future analysis of this variable, a follow up is integrated in the process of code co-development in order to extract use cases or success stories to make a qualitative assessment based on it.

Industry participation in PRACE events

The data available to PRACE project in this respect refers to the type of companies that have participated in the PRACE industrial seminars and the register of companies who has participated in the three workshops organized. Both indicators show a sustained interest of industry. The data gathered until now is probably not enough for reaching solid conclusions in terms of impact assessment, but a continuous monitoring of these values will provide a good indicator of the interests by industry in PRACE.

PRACE raising awareness events and media coverage

This variable has been assessed in terms of contacts generated in international events by PRACE and unique visits to the PRACE website since 2011. On the former, the data shows a significant baseline data on the number of interested people asking for information in PRACE booths. Further data will provide enough context to assess if the interest remains over time, decreases or increases. The data regarding web access shows an increasing trend in visits, which is quite natural because of the fact that PRACE is in its start-up process. In the analysis, we have differentiated two types of visitors: the scientists accessing to get information on how to get access to PRACE upon a public call is issued, and the visitors that want to be aware of the news. Once PRACE is consolidated beyond its initial period, the analysis of the general trend of number of visitors, as well as the increase in the number of non-scientist users will provide a straight indicator on the impact that PRACE is having in terms of awareness both at a scientific level and at a societal level.

Ecological imprint

The mission of PRACE is carried on at a cost of a certain ecological imprint. In order to create awareness of this and to work to reduce it, the energy consumption of the PRACE systems has been analysed. A quantitative way to assess how PRACE is impacting the environment has been created by mapping the energy consumption in terms of CO_2 generated. The calculated quantity of CO_2 per PFlop is 180g/Pflops. All the available PRACE systems together (8,05Pflops) produce every second the equivalent amount of the CO_2 emissions as a regular car covering 10km (assuming 145g/km). In a year time scale, the CO_2 produced by PRACE is equivalent to the CO_2 produced by 25000 cars travelling the earth diameter distance. It is advised to calculate the updated value upon changes in the technical elements in the infrastructure to assess the improvement of the ecological imprint, e.g. the upgrade of a machine or host facility.

4.2 Next Steps

Impact assessment is necessary in any research organisation. The organisation drivers have to be different from those considered by commercial and business oriented organizations that focus almost exclusively on how the actions impact revenues and costs reduction. Moreover, impact evaluation is addressed not to improve the operation of the infrastructure or the performance of plans, but to know the Return of Investment of the organisation, the

equivalent to the profit made for a profit-oriented organisation. But aside of the internal need to carry out this process, impact assessment of research infrastructures is becoming of key importance in the current downturn economic crisis to show to the society and to funding agencies what is the return of investment of science.

In Task 2.4 of PRACE-1IP, we have developed a theoretical framework for monitoring and reporting with a special emphasis in impact assessment. The current deliverable shows initial example analysis that should be carried out by the infrastructure or by future implementation assistance projects of the infrastructure.

Special emphasis should be on detailing the specific processes to integrate the impact assessment in the infrastructure taking into account the available resources, and also bring to discussion procedures to be implemented by computing centres to have a distributed network of data providers and analyses. Regarding reporting, a confidential channel for transmitting the data should be created from the data collectors to the organization. This means that the people involved in the process from collecting to receiving the data in the PRACE RI office should treat it confidentially signing if necessary non-disclosure agreements (NDAs).

Although the PRACE-1IP project has provided initial support to the Impact Assessment task developing the theoretical framework and baseline results, PRACE RI must take this task over in the near future, integrating the monitoring processes in its operation and its regular assessment for providing data to steer the organisation.