



Available online at www.prace-ri.eu

Partnership for Advanced Computing in Europe

An Administrative Framework for Tier-1 Resource Exchange within Europe

George Karagiannopoulos^a

Greek Research and Technology Network, 56, Mesogion Av. 11527, Athens, Greece

Ioannis Liabotis^b

Greek Research and Technology Network, 56, Mesogion Av. 11527, Athens, Greece

Wim Rijks^c

SURFsara, Science Park 140, 1098 XG Amsterdam, Netherlands

Abstract

DECI (Distributed European Computing Initiative) provides a resource exchange programme for Tier-1 resources within Europe. The programme has been running since 2005, initially under the DEISA projects and since September 2011 under the PRACE Implementation Phase projects. In PRACE IP projects, in contrast to DEISA, partners can opt in and out of calls on a call-by-call basis. This raises some issues in order to fulfill the *juste retour* principle of compute resource exchange between different partners. A stable administrative framework for resource exchange is necessary in order to provide PRACE with a tool to facilitate the management process of Tier-1 resource allocation during the DECI calls and to give a reliable mechanism to all partners on accounting of resource contribution and resource consumption. This White Paper describes the process, identifies the main actors, and presents the appropriate scheduling of the DECI Call preparation and implementation. We found that the main elements of such a framework are to record the resource commitments/availability for each DECI call and make those accessible to the relevant parties, to keep track of the usage of such resources and to provide views of provided vs consumed resources per DECI call, per year or for any arbitrary time period. The proposed framework could be useful for the organization committee in any implementation of a Programme for Tier-1 resources exchange within Europe, for the developers that will implement the proposed tool to support the DECI process and also for policy makers who will be responsible for the acquisition of new Tier-1 systems.

^a george.karagiannopoulos@grnet.gr

^b iliaboti@grnet.gr

^c Wim.Rijks@surfsara.nl

1. Introduction

DECI (Distributed European Computing Initiative) is a European single-project HPC access scheme which provides a resource exchange programme for Tier-1 (i.e. HPC systems under national governance) resources within Europe. DECI has been supported by the PRACE Implementation Phase projects since September 2011. Previously, DECI, which originally stood for DEISA Extreme Computing Initiative, was conceived of and supported by the DEISA projects which issued six DECI calls between 2004 and 2009.

Under the PRACE projects, DECI has been integrated into the PRACE ecosystem of HPC facilities and services. A main objective of continuing DECI in the PRACE implementation phase projects was to maintain the momentum gained by DEISA's DECI until an envisaged, more permanent Tier-1 programme is designed and implemented. In order to provide seamless access to users, PRACE project took over the organisation, operation and dissemination of DECI scientific results.

The challenge has been to expand the DECI programme and to maintain the high quality of project support, incorporating new partners promptly and bringing them to full effectiveness as quickly as possible. The number of sites providing Tier-1 access has grown steadily.

Participating DECI centres provide resources of around 5% of their national resource to DECI annually. More specifically, 85% of resources are pooled to be used for projects submitted by applicants from different countries than those contributing the resources and 15% are used for projects arriving from countries not contributing to the DECI resources pool. These resources are then distributed across successful DECI projects with the aim that projects within each country shall receive an amount of computing resources roughly equal (quantitatively) to the amount contributed but qualitatively different (i.e. on different classes of architecture).

In order to build a Tier-1 infrastructure, based on reciprocity of resources provided to PRACE Tier-1 resource exchange program by national HPC operators, a transparent model for resource evaluation needs to be agreed on and implemented and an administrative processes for evaluating, allocating and managing the resources needs to be designed in such a way that makes them easy to understand and operate. These will facilitate the addition of new HPC systems and reduce the overheads and costs of running a Tier-1 resource exchange programme.

The main objective is to formalize and add automation to the process of Tier-1 resource exchange management, minimizing the time and effort to collect and present the relevant information by creating various types of reports and fully avoiding possible errors that are associated with the current process.

2. DECI calls – analysis of resources

PRACE projects supported 5 DECI calls (DECI-7 pilot, DECI-8, DECI-9, DECI-10, DECI-11) since 2011. This means that two (2) calls per year have been tendered.

The following table presents data regarding each DECI call (except DECI-11 that is in the middle of the implementation).

Calls	No of Projects Awarded	No of contributing Countries	No of Sites	No of Machines	Committed Standard Core-h (DECI only)
DECI-7	35	11	15	17	90.771.356
DECI-8	33	11	15	19	100.575.324
DECI-9	31	12	14	14	129.455.315
DECI-10	37	16	17	19	168.247.919

3. Current implementation

The current situation is that each country declares its high level commitments of existing or future Tier-1 system resources that they will provide to the PRACE Tier-1 resource exchange programme. The data provided by each country includes the providing partner's name, the system's name, a brief and high level description of the system's architecture, the performance of the system using the LINPACK benchmark, the fraction of the system to be offered to PRACE for the duration of this DECI as well as the bandwidth of the dedicated connection to the PRACE network, where Tier-0 systems are also connected. It is worth mentioning that the earlier the information is collected the less reliable it is, since the future Tier-1 systems might still be at the procurement stage.

Before the official announcement of each Call, each DECI partner, who will participate in the call, provides its Tier-1 resource contribution. This information is collected and organized by the DECI co-ordinator on behalf of the Project Management Office (PMO) in a file (Excel) including fields such as site name, system's name and architecture, TFlop capacity from the LINPACK benchmark, number of cores, sites commitment as a percentage of the capacity, period over which this commitment is valid, including information on resources provided for the previous as well as the running call (% commitment, core hours, and several comments if deemed necessary).

This information is used by the DECI Team that converts those commitments to Standard core hours. Details on how to obtain these Standard core hours, which are equivalent to Power 4+ core hours, are provided in Deliverable 4.3.2 of PRACE-1IP [1].

After closing a particular call and finalizing the results of the accepted DECI projects and thus the upcoming application support requests, the system assignment and resource allocation takes place. This process is also described in the Deliverable 4.3.2 from PRACE-1IP. The data related to the above actions are included in a different Excel file.

4. DECI Project Management database

Data regarding the DECI projects (assigned machines, number of assigned core hours etc.) is inserted in the DECI Project Management Database (DPMDB). Several useful data views are provided in the DPMDB. These views are organized by project, execution site or home site.

The DECI database (DPMDB) has provided the mainstay of the DECI activity with all proposals having their data entered into this database together with links to their technical and scientific evaluations. Successful proposals are then turned into "real" projects in the DPMDB and more details, such as the machines to be used by each project, are added. During the course of a DECI project information is added to show the status of the project (enabling, performing production runs, completed, etc.) and the number of standard hours consumed by each project is entered. This makes it easy to obtain an overview of the status and progress of projects throughout their lifetime. Accounting data about used resources is inserted into the DPMDB on a monthly basis.

5. Administrative framework proposal

In order to setup a stable administrative framework for resource exchange, questions such as *who is doing what, when something has to be done, in which sequence, what information needs to be provided* need to be answered. A proposal for the technical implementation of such a resource exchange framework, via either a lightweight (wiki-based) procedure, or with the creation of a new web-based tool or via radical changes of the DPMDB, also needs to be presented.

We propose three (3) main actors in the Resource Exchange Framework, namely the Project Members or Management board of DECI (i.e. PRACE-IP management board or Tier-1 Resource Exchange Programme MB), hereafter referred to as **A1**, the Site/computing centre that hosts the system and the associated country who gives resources to the DECI Call, hereafter referred to as **A2**, and the DECI implementation team, hereafter referred to as **A3**, to whom **A1** has the authority to delegate tasks to.

We also present twelve (12) milestones (**M1-M12**) for the preparation and implementation of each DECI call regarding the Resource Exchange Framework. The Milestones of the workflow that require a tool based support are marked with – "new/ updated tool required", while others that can be supported by the existing infrastructure are marked with – "no new tool required".

The 1st milestone (**M1**) is the declaration of the partners' willingness to participate in the DECI program, so that their contribution to individual DECI calls can be verified whenever necessary (*there is a new/updated tool required*

for M1). M1 is executed by A1 by preparing a simple list of potential participants in the PRACE Implementation Project or the Tier-1 Resource Exchange Programme and high-level resource commitments.

The 2nd milestone (**M2**) is the initial declaration of resource availability for the next DECI call that needs to be in place, so that the feasibility of this call is evaluated (*there is a new/updated tool required for M2*). M2 needs to be actioned before the decision for the initiation of the call (Indicative best practice: finalization of resource availability one (1) month before the call) by A1 and A2. This is to be done by submitting and recording limited site information (such as Country, Partner, Site Name, Machine Name, LINKPACK [Tflop], % committed) by using the new /updated tool.

The 3rd milestone (**M3**) is for A1 to validate the next DECI Call date based on the pre-decided time-plan. Possible date changes might be decided during this process - (*no new tool is required for M3*). The validation should preferably take place one (1) month before the public announcement. It is understood that this time frame might not be possible to be followed in all cases.

The 4th milestone (**M4**) involves the insertion of the detailed information about the systems participating in the call by the site representatives (A2) using a *new/ updated tool* no later than one (1) month before the final call announcement. The information is the Country, Partner, Site Name, System Name, System Platform, CPU, CPU Factor (for Standard hour conversion), Interconnect, LINPACK [Tflops], number of cores, Total Memory, Available Storage, available from, comment, DECI call ID, % committed, committed core-hours, comment.

During the 5th milestone (**M5**) the above information is being validated and based on this, the call is being announced by A1 based on the DECI call schedule (*no new tool required*).

Milestone No 6 (**M6**) is the commencement of the proposal evaluation (*no new tool required*) by A3 based on DECI call schedule.

The 7th milestone (**M7**) involves the update and finalization of the information about available systems (see also M4) by A3 by submitting the final information (Country, Partner, Site Name, System Name, System Platform, CPU, CPU Factor, Interconnect, LINPACK [Tflops], no of cores, Total Memory, available from, comment, DECI call ID, % committed, committed core-hours, comment) by *using the new/ updated tool*. M7 should take place at the start of the Scientific Evaluation (SE) period (Based on DECI call schedule). Based on this up to date information the project to machine assignment is later performed (see M9)

The 8th milestone (**M8**) is when the final information about available systems needs to be imported in DPMDB from the new/ updated tool by A3 right before the assignment to machines (Based on DECI call schedule).

Milestone No 9 (**M9**) is about assigning the projects to machines (*no new tool required*). M9 is to take place after the finalization of the peer review by A3 with the assistance of project and machine information available in DPMDB.

The assigned projects start (*no new tool required*) based on DECI call schedule (10th milestone (**M10**)).

During the 11th milestone (**M11**) the accounting data shall be evaluated. This is to be done periodically (*a new/updated tool is required*) by inserting data about core hours consumed during the reporting period per site to DPMDB by sites (preferably in an automatic way).

Milestone No 12 (**M12**) is when the reports/ views will be produced upon request by *using the new/ updated tool* to compare committed vs used core hours for each system or partner over time periods (i.e. 1, 2, 3 years) or DECI calls and initial commitments vs actual commitments for each system or partner over time periods (i.e. 1, 2, 3 years) or DECI calls.

6. Assistance by a tool

In order for the framework presented above to be formalized we propose that support from a preferably web based tool is needed. The guidance of the process by a software tool will ensure efficient support of the coordination of the process, will minimize the errors, maximize the reliability of the information about the systems that will be used for a call, and ensure transparency and availability of information at any time to all stakeholders of the process. Due to the fact that radical changes to the DPMDB need significant effort and major security adjustments, the development/implementation of a new web-based tool (SIDB) that would be used for the registration of sites and systems that will participate to each DECI might be needed. This tool will consolidate the actions taken by the support of different existing mechanisms i.e. email, Excel, wiki. SIDB can either be implemented as a completely separate tool or a component of the DPMDB tool. The data from this tool will be imported to DPMDB without changing its current data structure. This would reduce the DPMDB admin effort to import site detailed configuration data and the errors that would happen when entering / updating the respective DPMDB information.

The new proposed tool has to support the creation and/or change of configuration data for access to the tool. i.e. user's, site's and country's info, the assignment of stakeholder's roles that can add the relevant information at different points in time (specific deadlines), the specification of milestones and time limits, the error control, the ability to import data from different applications/systems (i.e. wiki) and the capability to export data to different destinations (i.e. Excel, wiki etc).

As an intermediate solution the use of the PRACE-wiki is proposed. The use of the wiki as a tool for collecting information on Tier-1 systems has already started by the Operations activity of the PRACE-2IP project and at the moment it acts as an information point for registering and viewing information on the available systems and their characteristics for the DECI calls.

7. Resource Accounting

In order to gather accounting information, two systems can be used: DART (<http://www.prace-ri.eu/Accounting-Report-Tool>), or Grid-SAFE (<http://www.epcc.ed.ac.uk/projects/grid-safe>).

In principle the DPMDB has all the information needed in order to evaluate the resources offered from sites or countries to the DECI call. In future special views might need to be created for showing the information for different periods of time (i.e. 1, 2 and 3 years) and not only on a per DECI call basis. Site resource commitments are stored in the project's document store (BSCW), from where they are copied manually into DPMDB.

Through the above tools and mechanisms, DECI project management can evaluate resource commitments and actual usage of resources per partner or group of partners.

8. Discussion and conclusions

In this White Paper we describe the general principles of the proposed administrative framework for Tier-1 resource exchange framework within Europe. The main concerns are to formalize and to add automation to the process of Tier-1 resource exchange management. Minimizing the time and effort to collect the relevant information and avoiding the errors that are associated with the current process and collected data are also issues that would be resolved, as well as how to maximize the reliability of the information about the systems that will be used for a call.

Proposal submission and review process (technical and scientific evaluation) are out of the scope of the described framework, and are being implemented within the scope of the work done for the PPR (PRACE Peer Review DB) tool [4].

The proposed resource exchange framework provides a transparent view to all stakeholders about the lifecycle of the implementation of the DECI Calls i.e. available systems and computational resources [from the announcement of a call to the project allocation to sites and machines] in order to monitor the *juste retour* principle. It can also assist all DECI actors (Project Members or Management board of DECI, Sites/computing centres that host the systems and the associated countries who give resources to DECI Call, and the DECI organization team) to better organize their work and their obligations.

In this White Paper we give guidelines for the requirements of the implementation of a new tool or extension of existing tools to support all the concrete steps of the process. The implementation of such a tool to support the resource exchange framework is highly recommended to the actors that will design and implement the future DECI calls.

9. References

- [1] PRACE-1IP deliverable D4.3.2 "Cross-National Programme for Tier-1 Access Pilots"
- [2] PRACE-2IP deliverable D2.1 "Migration from DEISA2 to PRACE-2IP"
- [3] PRACE-2IP deliverable D2.1 "First Annual Report of WP2"
- [4] PRACE-2IP deliverable D10.1 "First Annual Report of WP10"

All deliverables are available at <http://www.prace-ri.eu/Public-Deliverables>

Acknowledgements

This work was financially supported by the PRACE project, funded in part by the EUs 7th Framework Programme (FP7/2007-2013) under grant agreement no. RI-283493.