TECHNICAL GUIDELINES FOR APPLICANTS TO PRACE 14th CALL (Tier-0)

Contributing sites and the corresponding computer systems for this call are:

GENCI – CEA, France	**Bull Bullx cluster "*Curie*"**
GCS – HLRS, Germany	**Cray XC40 System "*Hazel Hen*"**
GCS – JSC, Germany	**IBM BlueGene/Q System "*Juqueen*"**
CINECA, Italy	**Lenovo System "*Marconi*"**
BSC, Spain	**IBM System X iDataPlex "*MareNostrum*"**
CSCS, Switzerland	**Cray XC30 System "*Piz Daint*"**
GCS – LRZ, Germany	**IBM System X iDataPlex "*SuperMUC*"**

The site selection is done together with the specification of the requested computing time by the two sections at the beginning of the online form. The applicant can choose one or several machines as execution system.

The parameters are listed in tables. The first column describes the field in the web online form to be filled in by the applicant. The remaining columns specify the range limits for each system.

The applicant should indicate the unit.

A - General Information on the Tier-0 systems available for PRACE 14th Call

		Curie TN	***Hazel Hen***	***Juqueen***	***Marconi Broadwell***	***Marconi KNL***	***MareNostrum***	***Piz Daint***	***SuperMUC Phase 1***	***SuperMUC Phase 2***
System Type		Bullx	Cray XC40	Blue Gene/Q	Lenovo System NeXtScale	Lenovo System Adam Pass	IBM System x iDataPlex	Hybrid Cray xC30	IBM System x iDataPlex	Lenovo NeXtScale
Compute	Processor type	Intel SandyBridge EP 2.7 GHz	Intel Xeon E5-2680v3 (Haswell)	IBM PowerPC® A2 1.6 GHz 16 cores per node	Intel Broadwell	Intel Knights Landing	Intel Sandy Bridge EP	SandyBridge Upgrade to Haswell starting Oct 17	Intel Sandy Bridge EP	Haswell Xeon E5-2697 v3 (Haswell)
	Total nb of nodes	5 040	7 712	28 672	1 512	3 600	3 056	5 272	9 216	3 072
	Total nb of cores	80 640	185 088	458 752	54 432	244 800	48 896	84 352 (8x2)	147 456	86 016
	Nb of accelerators/node	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	5272	n.a.	n.a.
	Type of accelerator	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	Kepler K20X Upgrade to Pascal strating Oct 17	n.a.	n.a.
Memory	Memory / Node	64 GB	128 GB	16 GB	128 GB - DDR4	96 GB – DDR4 + 16 GB - MCDRAM	32 GB	32 GB	32 GB	64 GB
Network	Network Type	Infiniband QDR	Cray Aries	5D Torus network	Intel Omni-Path Architecture 2:1	Intel Omni-Path Architecture 2:1	Infiniband FDR10	Cray Aries	Infiniband FDR10	Infiniband FDR14
	Connectivity	Fat tree	Dragonfly	5D Torus network	Fat Tree	Fat Tree	Fat Tree	Dragonfly	Fat tree within island (512 nodes) pruned tree between islands	Fat tree within island (512 nodes) pruned tree between islands

* Only for pre-/post-processing purposes. Please verify with Gauss@LRZ for large computing resource requests on FN.

		Curie	*Hazel Hen*	*Juqueen*	*Marconi Broadwell*	*Marconi KNL*	*MareNostrum*	*Piz Daint*	*SuperMUC Phase 1*	*SuperMUC Phase 2*
Home file system	type	NFS	NFS	GPFS	GPFS	GPFS	GPFS	GPFS	NAS *	NAS *
	capacity	8 TB	60 TB	1.8 PB	200 TB	200 TB	59 TB	80 TB	10 PB	10 PB
Work file system	type	Lustre	Lustre	n.a.	GPFS	GPFS	GPFS	GPFS	GPFS *	GPFS *
	capacity	600 TB	12 PB	n.a.	7.1 PB	7.1 PB	612 TB	5.7 PB	12 PB	12 PB
Scratch file system	type	Lustre	n.a.	GPFS	GPFS	GPFS	GPFS	Lustre	GPFS *	GPFS *
	capacity	3.4 PB	n.a.	5.3 PB	2.5 PB	2.5 PB	1.1 PB	2.7 PB (to be upgraded with the system)	5.2 PB *	5.2 PB *
Archive	capacity	Unlimited	On demand	Unlimited	On demand	On demand	3.7 PB	NA	On demand	On demand
Minimum required job size	Nb of cores	1 024	2 048	2 048	540	2 040	1 024	1 024	512	512

* SuperMUC Phase 1 and Phase 2 share the same file systems

IMPORTANT REMARK:

Further information and support from HPC Technical teams can be requested through PRACE Peer-Review at peer-review@prace-ri.eu during the preparation of the application.

MARCONI:

Since MARCONI will be the first system in Europe equipped with KNL processors and Intel Omni-Path network, it is recommended to applicants to prove that their application codes are able to scale at least on similar Intel architectures, interconnected at least with Infiniband like fat tree topology. This will be favourably considered for the technical evaluation of the project.

MareNostrum:

MareNostrum III will be replaced by a new system named MareNostrum IV. All applications running in MareNostrum III will also be supported on the newer system and for now the terms of access will be the same as for the existing system.

More details on the website of the centers:

Curie:
http://www-hpc.cea.fr/en/complexe/tgcc-curie.htm

Hazel Hen:
http://www.hlrs.de/systems/cray-xc40-hazel-hen/

Juqueen:
http://www.fz-juelich.de/ias/jsc/EN/Expertise/Supercomputers/JUQUEEN/JUQUEEN_node.html

Marconi:
http://www.hpc.cineca.it/hardware/marcon
https://wiki.u-gov.it/confluence/display/SCAIUS/UG3.1%3A+MARCONI+UserGuide

MareNostrum:
http://www.bsc.es/marenostrum-support-services/mn3

Piz Daint:
http://user.cscs.ch/index.html

SuperMUC:
http://www.lrz.de/services/compute/supermuc/

Subsection for each system

Curie, GENCI@CEA

The Curie BULLx system is composed by 5040 compute blades, each node having 2 octo-core Intel SandyBridge EP processors 2.7 GHz, 4 GB/core (64 GB/node) and around 64 GB of local SSD acting as local /tmp. These nodes are interconnected through an Infiniband QDR network and accessing to a multi-layer Lustre parallel filesystem at 250 GB/s. The peak performance of system is 1.7 petaflops.

Hazel Hen, GCS@HLRS

Hazel Hen, a Cray XC40-system, is at the heart of the high performance computing (HPC) system infrastructure of the HLRS. With a peak performance of 7.42 Petaflops (quadrillion floating point operations per second), Hazel Hen is one of the most powerful HPC systems in the world (position 8 of TOP500, 11/2015) and is the fastest supercomputer in the European Union. The HLRS supercomputer, which was taken into operation in October 2015, is based on the Intel® Haswell Processor and the Cray Aries network and is designed for sustained application performance and high scalability.

Juqueen, GCS@JSC

The Blue Gene/Q system JUQUEEN consists of 28 Racks (28 672 compute nodes). One node consists of 16 cores with four-fold SMT and is equipped with 16 GB of main memory.

Typically about 15 GB of main memory per compute node are available for user applications. This means each physical core has a bit less than 1 GB of main memory available.

In order to use the Blue Gene/Q architecture efficiently, pure MPI codes must use 32 tasks per node, hybrid codes must use the corresponding number of threads per task (from 32 to 64 threads per node in total).

Therefore, pure MPI applications must use less than 512 MB per MPI task, hybrid codes less than 16 GB per node in order to be suited for the architecture.

Marconi, CINECA

Marconi system consists of three partitions, from which two will be available for Call 14:

- **Marconi - Broadwell** consists of ~21 Lenovo NeXtScale racks with 72 nodes per rack. Each node contains 2 Broadwell processors each with 18 cores and 128 GB of DDR4 RAM.
- **Marconi - KNL** will be deployed by the end of 2 016 and will consist of 3 600 Intel server nodes integrated by Lenovo. Each node contains 1 Intel Knights Landing processor with 68 cores, 16 GB of MCDRAM and 96 GB of DDR4 RAM.

All the nodes (for both systems) will be connected via Intel Omni-Path network.

MareNostrum, BSC

- **MareNostrum system** consists of 36 IBM iDataPlex Compute Racks, and 84 IBM dx360 M4 compute nodes per rack. Each compute node has two 8-core SandyBridge-EP processors at 2.6 GHz, and 32 GB of main memory (2 GB/core), connected via Infiniband FDR10.

MareNostrum will be upgraded on the mid-2017, this system will be a general purpose processor with a minimum of 2GB/core and a high bandwidth and low latency network connection with up to 100Gbits/s. It will be a minimum of 9 Pflops/s and will be able to run all the current codes available for the previous MareNostrum machine.

Piz Daint, CSCS

Named after Piz Daint, a prominent peak in Grisons that overlooks the Fuorn pass, this supercomputer is a hybrid Cray XC30 system and is the flagship system for national HPC Service. This supercomputer is a 28 cabinet Cray XC30 system with a total of 5 272 compute nodes. The compute nodes are equipped with an 8-core 64-bit Intel SandyBridge CPU (Intel® Xeon® E5-2670), an NVIDIA® Tesla® K20X with 6 GB GDDR5 memory, and 32 GB of host memory.

The nodes are connected by the "Aries" proprietary interconnect from Cray, with a dragonfly network topology.

Piz Daint is going to be upgraded from SandyBridge(CPU)/K20x(GPU) to a Haswell(CPU)/Pascal(GPU) starting on October 17, 2016

SuperMUC, GCS@LRZ

SuperMUC Phase 1 consists of 18 Thin Node Islands with Intel Sandy Bridge processors and one Fat Node Island with Intel Westmere processors. Each Island contains slightly more than 8192 cores. Each of these cores has approx. 1.5GB/core available for running applications. Peak performance is 3.1 PF. All compute nodes within an individual Island are connected via a fully non-blocking Infiniband network (FDR10 for the Thin Nodes and QDR for the Fat Nodes). A pruned tree network connects the Islands.

SuperMUC Phase 2 consists of 6 Islands based on Intel Haswell-EP processor technology (512 nodes/island, 28 physical cores/node and available memory 2.0 GB/core for applications, 3 072 nodes, 3.6 PF). All compute nodes within an individual Island are connected via a fully non-blocking Infiniband network (FDR14). A pruned tree network connects the Islands.

Both system phases share the same Parallel and Home filesystems.

B – Guidelines for filling-in the on-line form

Resource Usage

Computing time

The amount of computing time has to be specified in core-hours (wall clock time [hours]*physical cores of the machine applied for). It is the total number of core-hours to be consumed within the twelve months period of the project.

Please justify the number of core-hours you request with **a detailed work plan**. Not doing so might result in decreasing the amount of core-hours or even in rejection of the proposal.

The project should be able to start immediately and is expected to use the resources continuously.

When planning for access, please take into consideration that the effective availability of the system is about 80 % of the total availability, due to queue times, possible system maintenance, upgrade, and data transfer time.

If less than 15 million core-hours (see Terms of Reference) in one of the Tier-0 system is required, the choice to use Tier-0 systems has to be justified as compared to the use of Tier-1 systems.

The maximum value of computing time is limited by the total number of core-hours per system given in the terms of reference document for the 14th Call (see the “Call announcement” page at www.prace-ri.eu/Call-Announcements). **Any further limitation is specified in the terms of reference document of the corresponding Call for Proposals**.

Job Characteristics

This section describes technical specifications of simulation runs performed within the project.

Wall Clock Time

A simulation consists in general of several jobs. The wall clock time for a simulation is the total time needed to perform such a sequence of jobs. This time could be very large and could exceed the job wall clock time limits on the machine. **In that case the application has to be able to write checkpoints and the maximum time between two checkpoints has to be less than the wall clock time limit on the specified machine.**

Field in online form	*Machine*	*Max*
Wall clock time of one typical simulation (hours) **<number>**	Other systems Piz Daint	< 10 months 24 hrs
Able to write checkpoints **<check button>**	All	Yes
Maximum time between two checkpoints **(= maximum wall clock time for a job) (hours)** **<number>**	Other systems Hazel Hen Juqueen Piz Daint SuperMUC	24 hours 24 hours (12 hours)* 24 hours (12 hours)* 24 hours 48 hours

* This might be changed during project runtime, guaranteed minimum is the value in brackets.

Number of simultaneously running jobs

The next field specifies the number of independent runs which could run simultaneously on the system during normal production conditions. This information is needed for batch system usage planning and to verify if the proposed work plan is feasible during project run time.

Field in online form	*Machine*	*Max*
Number of jobs that can run simultaneously <number>	Curie Hazel Hen Juqueen Marconi MareNostrum Piz Daint SuperMUC	25 (1 024 cores), 4 (8 192 cores) 29 and at maximum 60.000 cores all jobs together 15 2-10 (depending on the job size) Dynamic* We do not have shared nodes 1 job per node maximum 10 (512 cores), 3 (8 192 cores), 1 (>32 768 cores)

* Depending on the amount of PRACE projects assigned to the machine, this value could be changed.

Job Size

The next fields describe the job resource requirements which are the number of cores and the amount of main memory. These numbers have to be defined for three different job classes (with minimum, average, or maximum number of cores).

Please note that the values stated in the table below are absolute minimum requirements, allowed for small jobs, which should only be requested for a small share of the requested computing time. Typical production jobs should run at larger scale.

Job sizes must be a multiple of the minimum number of cores in order to make efficient use of the architecture.

IMPORTANT REMARK

Please provide explicit scaling data of the codes you plan to work with in your project at least up to the minimum number of physical cores required by the specified site (see table below) using input parameters comparable to the ones you will use in your project (a link to external websites, just referencing other sources or "general knowledge" is not sufficient). ***Generic scaling plots provided by vendors or developers do not necessarily reflect the actual code behavior for the simulations planned. Missing scaling data may result in rejection of the proposal.***

Field in online form	*Machine*	*Min (cores)*
Expected job configuration (<u>Minimum</u>) <number>	Curie Hazel Hen Juqueen Marconi MareNostrum Piz Daint SuperMUC	1 024 2 048 2 048 540 (Broadwell) 2 040 (KNL) 1 024 1 024 512
Expected number of cores (<u>Average</u>) <number>	Curie Hazel Hen Juqueen Marconi MareNostrum Piz Daint SuperMUC	4 096 8 192 4 096 1 080 (Broadwell), 4 080 (KNL) 1 024 1 024 minimum + 1 GPU per node 16 384 (Phase1), 7 168 (Phase2)
Expected number of cores (<u>Maximum</u>) <number>	Curie Hazel Hen Juqueen Marconi MareNostrum Piz Daint SuperMUC	40 000 (80 000 on demand) 96 000 (using up to 113 664 cores is possible, but should be requested in the proposal) > 4 096 (up to the complete system) 6 000 (Broadwell), 68 000 (KNL) n.a. 42 176 (84 352 in hyp-thr) 65 536 (Phase1), 14 426 (Phase2)

Virtual cores (SMT is enabled) are not counted. Accelerator based systems (GPU, Xeo,, Phi, etc) need *special rules*.

<u>Additional information:</u>

<u>Marconi</u>

The minimum number of (physical) cores per job is 2 040 (only for KNL partition. 540 on Broadwell).

However, this minimum requirement should only be requested for a small share of the requested computing time and it is expected that PRACE projects applying for Marconi can use at least a doubled value on average.

The maximum number of (physical) cores per job is 68 000. Larger jobs are possible in theory (only if requested sending an email to superc@cineca.it), but the turnaround time is not guaranteed.

Please provide explicit scaling data of the codes you plan to work within your project. On KNL partition a good scalability up to 4 000 physical cores must be demonstrated and the scaling behavior up to 8000 physical cores must be shown using input parameters comparable to the ones you will use in your project.

For hybrid (MPI+OpenMP) codes it is strongly recommended that applicants show scaling data for different numbers of threads per task in order to exploit the machine most efficiently. Providing such kind of data will be favorably considered for the technical evaluation of the project.

Since MARCONI – KNL will be one of the first systems that uses Knights Landing Processors, only for this call we will accept scalability test performed on other Intel architectures (e.g.: Sandy Bridge, Ivy Bridge, Haswell, Knights Corner).

Proving to be able to scale on other architectures will be favorably considered for the technical evaluation of the project.

<u>JUQUEEN</u>

The minimum number of (physical) cores per job is 2048. However, this minimum requirement should only be requested for a small share of the requested computing time and it is expected that PRACE projects applying for JUQUEEN can use more than 2048 physical cores per job on average (at least 4096 physical cores). Job sizes must use a multiple of 2048 physical cores in order to make efficient use of the architecture.

Please provide explicit scaling data of the codes you plan to work with in your project. A good scalability up to 4096 physical cores must be demonstrated and the scaling behavior up to 8192 physical cores must be shown using input parameters comparable to the ones you will use in your project.

For hybrid (multi-threaded) codes it is strongly recommended, that applicants show scaling data for different numbers of threads per task in order to exploit the machine most efficiently. Providing such kind of data will be favorably considered for the technical evaluation of the project.

SuperMUC

The minimum number of (physical) cores per job is 512. However, it is expected that PRACE projects applying for this system can use more than 2048 physical cores per job. When running several jobs simultaneously filling complete islands should be possible but this is not mandatory.

Job Memory

The next fields are the total memory usage over all cores of jobs.

Field in online form	***Machine***	***Max***
Memory (Minimum job) **<number>**	Curie Hazel Hen Juqueen Marconi MareNostrum Piz Daint SuperMUC	- 4 GB * #cores or 64 GB * #nodes - Jobs should use a substantial fraction of the available memory - 1 GB * #cores - <118 GB per node (Broadwell), <90 GB per node (KNL) - 2 GB * #cores - n.a. - Jobs should use a substantial fraction of the available memory
Memory (Average job) **<number>**	Curie Hazel Hen Juqueen Marconi MareNostrum Piz Daint SuperMUC	- 4 GB * #cores or 64 GB * #nodes - Jobs should use a substantial fraction of the available memory - 1 GB * #cores - <118 GB per node (Broadwell); <90 GB per node (KNL) - 2 GB * #cores - n.a. - Jobs should use a substantial fraction of the available memory
Memory (Maximum job) **<number>**	Curie Hazel Hen Juqueen Marconi MareNostrum Piz Daint SuperMUC	- 4 GB* #cores or 64 GB * #nodes - 5GB* #cores - 1GB* #cores for the other cores - 1 GB* #cores - <118 GB per node (Broadwell), <90 GB per node (KNL) - 2 GB* #cores - n.a. - 1 5 GB* #cores or 24 GB* #nodes (Phase1) - 2.0 GB* #cores or 56 GB* #nodes (Phase2)

The memory values include the resources needed for the operating system, i.e. the application has less memory available than specified in the table.

Storage

General remarks

The storage requirements have to be defined for four different storage classes (Scratch, Work, Home and Archive).

- **Scratch** acts as a temporary storage location (job input/output, scratch files during computation, checkpoint/restart files; no backup; automatic remove of old files).
- **Work** acts as project storage (large results files, no backup).
- **Home** acts as repository for source code, binaries, libraries and applications with small size and I/O demands (source code, scientific results, important restart files; has a backup).
- **Archive** acts as a long-term storage location, typically data reside on tapes. For PRACE projects also archive data have to be removed after project end. The storage can only be used to backup data (simulation results) during project's lifetime.

Data in the archive is stored on tapes. **Do not store thousands of small files in the archive, use container formats** (e.g. tar) to merge files (**ideal size of files: 500 – 1 000 GB**). Otherwise, **you will not be able to retrieve back the files from the archive within an acceptable period of time** (for retrieving one file about 2 minutes time (independent of the file size!) + transfer time (dependent of file size) are needed)!

IMPORTANT REMARK

All data must be removed from the execution system within 2 months after the end of the project.

Total Storage

The value asked for is the maximum amount of data needed at a time. Typically this value varies over the project duration of 12 month (or yearly basis for multi-year projects). **The number in brackets in the "Max per project" column is an extended limit, which is only valid if the project applicant contacted the center beforehand for approval**.

Field in online form	***Machine***	***Max per project***	***Remarks***
Total storage (Scratch) <number> Typical use: Scratch files during simulation, log files, checkpoints Lifetime: Duration of jobs and between jobs	Curie	20 TB (100 TB)	- without backup, automatic clean-up procedure
	Hazel Hen	-	- HLRS provides a special mechanism for Work spaces, without backup
	Juqueen	20 TB (100 TB)	- without backup, files older than 90 days will be removed automatically
	Marconi	20 TB (100 TB)[*1]	- without backup, clean-up procedure for files older than 30 days
	MareNostrum	100 TB (200 TB)	- without backup, clean-up procedure
	Piz Daint	2.7 PB	- 24 hrs, without backup, clean-up procedure
	SuperMUC	100 TB (200 TB)	- without backup, automatic clean-up procedure
Total storage (Work) <number> Typical use: Result and large input files Lifetime: Duration of project	Curie	1 TB	-
	Hazel Hen	250 TB	- *2
	Juqueen	n.a.	- Not available on JUQUEEN
	Marconi	20 TB (100 TB)[*1]	- Without backup
	MareNostrum	10 TB	- With backup
	Piz Daint	5.6 PB	- Input not readable from compute nodes
	SuperMUC	100 TB (200 TB)	- Without backup

Total storage (<u>Home</u>) **<number>** **Typical use: Source code and scripts** **Lifetime: Duration of project**	Curie Hazel Hen Juqueen Marconi MareNostrum Piz Daint SuperMUC	3 GB 50 GB *3 6 TB 50 GB 40 GB 60 TB 100 GB	- with backup and snapshots - no backup - with backup - - with backup - - with backup and snapshots
Total storage (<u>Archive</u>) **<number>**	Curie Hazel Hen Juqueen Marconi MareNostrum Piz Daint SuperMUC	100 TB *4 *5 20 TB (100 TB)*6 100 TB n.a. 100 TB*7	- File size > 1 GB - - Ideal file size: 500 GB – 1000 GB - - - - Typical file size should be > 5 GB

*1 The default value is 1 TB. Please ask to CINECA User Support (superc@cineca.it) to increase your quota after the project will start

*2 The number given depends also on the number of users in the project.

*3 The number given depends also on the number of users in the project.

*4 Access to Hazel Hen's archive needs a special agreement with HLRS and PRACE.

*5 Due to limited file system cache for archive not more than 10 TB/week should be moved to this storage.

*6 Not active by default. Please ask to CINECA User Support after the project will start

*7 Long-term archiving or larger capacity must be negotiated separately with LRZ.

When requesting more than the specified scratch disk space and/or larger than 1 TB a day and/or storage of more than 4 million files, please justify this amount and describe your strategy concerning the handling of data (pre/post processing, transfer of data to/from the production system, retrieving relevant data for long-term). <u>If no justification is given the project will be proposed for rejection.</u>

If you request more than 100 TB of disk space, please contact peer-review@prace-ri.eu before submitting your proposal in order to check whether this can be realized.

Number of Files

In addition to the specification of the amount of data, the number of files also has to be specified. If you need to store more files, **<u>the project applicant must contact the center beforehand for approval</u>**.

Field in online form	***Machine***	***Max***	***Remarks***
Number of files (<u>Scratch</u>) **<number>**	Curie Hazel Hen Juqueen Marconi MareNostrum Piz Daint SuperMUC	2 Milion n.a. 4 Million 2 Milion 2 Milion 100.000 1 Million	- - Without backup files older than 90 days will be removed automatically - - No limit while running, but we have limits # number of files left on scratch - Without backup, old files are removed automatically, Ideal file size: >100 GB
Number of files (<u>Work</u>) **<number>**	Curie Hazel Hen Juqueen Marconi MareNostrum Piz Daint SuperMUC	500 000 4 Million n.a. 2 Milion 2 Milion 250 000 1 Million	- Extensible on demand - - - - - - Ideal file size: >100 GB

Number of files (Home) **<number>**	Curie Hazel Hen Juqueen Marconi MareNostrum Piz Daint SuperMUC	n.a. 100.000 12 Million 100 000 100 000 100 000 100 000	- - - With backup - - - - With backup(snapshots)
Number of files (Archive) **<number>**	Curie Hazel Hen Juqueen Marconi MareNostrum Piz Daint SuperMUC	100 000 10 000 2 Million 10 000* 1 Million* n.a. 100 000	- Extensible on demand - - Ideal file size: 500 GB – 1000 GB - - - - Typical file size should be > 5 GB

*HSM has a better performance with a small amount of very big files

Data Transfer

For planning network capacities, applicants have to specify the amount of data which will be transferred from the machine to another location. Field values can be given in Tbyte or Gbyte.

Reference values are given in the following table. *A detailed specification would be desirable: e.g. distinguish between home location and other PRACE Tier-0 sites.*

Please state clearly in your proposal the amount of data which needs to be transferred after the end of your project to your local system. Missing information may lead to rejection of the proposal.

Be aware that transfer of large amounts of data (e.g. tens of TB or more) may be challenging or even unfeasible due to limitations in bandwidth and time. Larger amounts of data have to be transferred continuously during project's lifetime.

Alternative strategies for transferring larger amounts of data at the end of projects have to be proposed by users (e.g. providing tapes or other solutions) and arranged with the technical staff.

Field in online form	***Machine***	***Max***
Amount of data transferred to/from production system **<number>**	Curie Hazel Hen Juqueen Marconi MareNostrum Piz Daint SuperMUC	100 TB 100 TB* 100 TB 20 TB* 50 TB Currently no limit 100 TB

* More is possible, but needs to be discussed with the site prior to proposal submission.

If one or more specifications above is larger than a reasonable size (e.g. more than tens of TB data or more than 1TB a day) the applicants must describe their strategy concerning the handling of data in a separate field (pre/post-processing, transfer of data to/from the production system, retrieving relevant data for long-term). In such a case, the application is *de facto* considered as I/O intensive.

I/O

Parallel I/O is mandatory for applications running on Tier-0 systems. Therefore the applicant must describe how parallel I/O is implemented (checkpoint handling, usage of I/O libraries, MPI I/O, netcdf, HDF5 or other approaches). Also the typical I/O load of a production job should be quantified (I/O data traffic/hour, number of files generated per hour).