



## **The “Jonker Case” / After Care: Handling the “ENTRAIN” Dataset after its Production on Jugene**

Huub Stoffers

(SARA)

---

### **Abstract**

“Huygens”, the IBM P6 system in Amsterdam and current incarnation of the Dutch National supercomputer for the academic community, was one of the DEISA systems and is now a PRACE Tier-1 system in the PRACE 2IP project. In the PRACE preparatory phase it also was a prototype system. An informal “PRACE\_HOME” storage space on Huygens provides a “next stop” for PRACE Tier-0 produced data that have to be preserved for a longer time. The experience feedback presented is tied with the project of a Dutch investigator, Harm Jonker. The simulation produces an important amount of output data to be preserved, and some issues were encountered with the data preservation.

---

### **1. File system and storage resources**

“Huygens”, the IBM P6 system in Amsterdam and current incarnation of the Dutch National supercomputer for the academic community, was one of the DEISA systems and is now a PRACE Tier-1 system in the PRACE-2IP project. In the PRACE preparatory phase it also was a prototype system.

In January 2011 we dismantled a more or less isolated environment, which had been used in the preparatory phase for tests that were considered too disruptive or have too much impact on the performance of the standard Huygens production environment. The isolated environment was built partly from P6-575 compute and InfiniBand interconnect resources, donated by IBM for the purpose, and partly from storage capacity that was taken away from Huygens’ home directory environment - temporarily, i.e.: originally the plan had been to return the storage to the home directory capacity and enlarge the default user quota there. In a SAN maintenance session on January 24th 2011 the storage resources of this isolated test setup were returned to Huygens environment where they subsequently were used however, to create the NEW file system with a net capacity of approximately 130 TB and capable of a sustained write performance of more than 2 GB/sec. and a read performance approaching 3 GB/sec. The file system is a GPFS file system directly accessible by all Huygens nodes. It presently is not, but could also be exported to other sites connected in the DEISA network and GPFS infrastructure.

This file system can be used as a “PRACE\_DATA” space, loosely analogous to the “DEISA\_DATA” space which, in a DEISA context, is provided by the “home site” and is available to projects, irrespective of the site where the project runs. First experiences with a PRACE project and problems it faced with the longer term preservation of data, and gradually becoming more involved with the practical implications of the subject matter to be addressed, convinced us that Huygens, as a PRACE Tier-1 site, will need something of the sort in PRACE too, if it is supposed to be one of the alternatives for longer term preservation of data produced on a Tier-0 site. SARA has a petabyte scale tape archiving facility which is connected to Huygens, and to some other HPC systems at SARA, but which is not directly accessible over the Internet by external systems. What is needed to make effective use of the archive is some sort of project staging/buffering space – a file system with substantial performance and multiple terabytes available for projects on a system that has ample network resources to engage in network data transfers with peers. At SARA, Huygens is that system.

### **2. First experiences with a PRACE project**

The “first experiences” referred to above are with the project of a Dutch investigator, Harm Jonker, of Delft Technical University in The Netherlands, titled: "Providing fundamental laws for weather and climate models".

The project has been accepted by PRACE and was granted 35,000,000 core hours on the PRACE Tier-0 resource JUGENE, the IBM Blue Gene/P at FZJ Jülich, Germany. The investigator and this project are clearly part of the community, c.q. field, of climate studies that has been selected by PRACE- IIP as one of the areas of particular interest.

The project proposal explicitly states that the data that are produced by "this project will be made publicly available so as to serve as benchmark for atmospheric models". However, the volume of the project's output is not stated explicitly, and no resources for longer term data storage are explicitly requested in the project proposal. Nor has the PRACE project appraisal procedure given any attention to this detail. Meanwhile most of the granted core hours on the Tier-0 resources have been used. The simulation runs on the JUGENE system have been completed successfully and produced 13 TB of output data to be preserved.

At this point it became rather manifest that something was lacking to make the project truly useful. As far as PRACE was concerned, the project was completed successfully: the granted core hours had been used and had generated the desired output. The output now merely had to vanish from the Tier-0 system to make room for newly scheduled projects. At this stage we noticed the investigator trying to "archive" his 13 TB of output data from Jülich, to a not particularly well suited<sup>a</sup> file system on Huygens, which was not intended for this purpose at all, using a rather inefficient transfer protocol.

The reason for trying to help out this particular investigator was of course not that he simply had happened to target Huygens - where he happened to have another account related to a different project - while "scavenging" for storage for his Tier-0 produced data. In the communication that followed it became fairly clear that the Tier-0 produced data would have to be post-processed and analyzed, and that this would lead to follow up proposals, that would typically require compute resources of a scale that is associated with a Tier-1 facility rather than with a follow up Tier-0 project. Given the background of the investigator, NCF seems to be the logical funding agency for such follow up projects, and hence Huygens, the present Dutch national supercomputing facility and the associated tape archive infrastructure at SARA, indeed seem to be the "natural" first choice for preserving these data.

### 3. PRACE Tier-1 sites and "Home Sites"

Much is ad hoc improvised and "taken for granted", in want of more explicit responsibilities and procedures. At first sight the PRACE procedures for applying for Tier-0 resources may look very similar to the DECI projects in the DEISA context, where users from one country are granted resources on computer systems of a different country too. However, in DEISA users are entered in an LDAP-based common user administration by the partner that is identified as "the home site" and the home site is responsible for contributing a DEISA\_HOME and a DEISA\_DATA storage area, irrespective of where the compute cycles are actually used. The DEISA\_HOME and DEISA\_DATA storage area are usually accessible by the involved sites by means of shared file system technology. But even if this is not the case, the fact that they are allocated and coupled to a DEISA user ID when the project starts, implies that they can be used legitimately and fairly efficiently as the destination in file transfers – as a "portal" to archive facilities present at the home site that has taken responsibility for this user.

The above is not to suggest that PRACE cannot do without a shared file system. It can. It is neither that there should be a PRACE\_HOME and/or PRACE\_DATA that are perfect analogies of their DEISA counterparts. But the fact that these exist in the DEISA setup ensures that a default "next stop" for data produced is already available for. What is clearly lacking in PRACE so far is support and guidance for users that goes beyond the stage that the allocated compute cycles on the Tier-0 resource have been used. From a "best practice" point of view the following topics should be explicitly addressed, by the applying investigators and by the PRACE procedural framework:

- A SUITABLE DESTINATION should be explicitly identified, especially for project proposals that explicitly state that they will generate output that has to be preserved for a specific period.
- There should be an assessment of HOW MUCH data need to be transferred to such a destination.
- An inquiry into the MEANS of transferring such data (network infrastructure, protocols) efficiently should also be part of the of the project's technical review,

---

<sup>a</sup> User home directory quotas on Huygens are far from sufficient to store 13 TB of data. There is a multi terabyte "scratch space", which is intended, used, and indeed managed as scratch space. Just to name one problematic aspect: data coming in from external sources over file transfer protocols that preserve file modification times on the destination side are almost immediately swept away by automated cleaning procedures that ensure that "old data" are not lingering too long (i.e.: longer than 14 days) and taking away room intended for running jobs.

- Furthermore, if it is clear from the outset that post-processing on the data is needed, PRACE project appraisal procedures should also inquire whether the local or “private” resources of the investigator are sufficient. If not, projects should make convincingly clear, that adequate arrangements have been made to ensure follow up resources.

The concept of a “home site”, borrowed from DEISA, however, is foreign to the PRACE setup. There is the issue of integration of Tier-0 sites with Tier-1 sites to be dealt with in PRACE-2IP and the work plan mentions the Tier-1 sites as a possible resource for long term preservation of data. But as far as I know the concept of “PRACE Tier-1 site” is not very well defined in terms of operational responsibilities and procedures.

Adoption of the home site concept by PRACE in the above outlined sense would be one more decentralized solution or “good practice” for handling longer term preservation of Tier-0 produced output. Checking this in the project appraisal phase ensures that Tier-1 sites are involved at an early stage and commit themselves to providing solutions for projects of which are now somehow rather vaguely expected or assumed to handle. It is of course not the only solution. It can be complementary to a conceivably more centralized practice in which PRACE Tier-0 sites setup specific archive facilities that are also well reachable by Tier-1 sites that investigators have in mind for follow up work. At Jülich there currently is no PRACE archive that meets these criteria. So what we have in fact done with the recent SAN and file system restructuring at Huygens is facilitating that Huygens as a Tier-1 site de facto CAN assume some of the responsibilities that “home sites” have in DEISA. What is subsequently still lacking is something more structural than mere “ad hoc” decisions taken after a problem has risen, to play such a role for a specific project.

#### **4. 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-211528 and FP7-261557.