SEVENTH FRAMEWORK PROGRAMME
Research Infrastructures

INFRA-2007-3.3: "Studies, conferences and coordination actions supporting policy development, including international cooperation, for e-Infrastructures"

LinkSCEEM

Linking Scientific Computing in Europe and the Eastern Mediterranean

Grant Agreement Number: 222904

D3
Final Report on Assessment of Needs

Version: 1.2

Contents

Executive Summary 1

1. Introduction 2

2. Users Database 3

2.1. Database as of January 2009 3

2.2. Current Database 6

3. User surveys 7

3.1. Simplified survey: 8

3.2. Detailed survey 15

3.3. Conference Survey 20

3.4. 3 Day HPC Workshop Survey 25

4. Panel Discussions of User Meetings 27

4.1. First Users Meeting of the LinkSCEEM Project, 8th April 2008 28

4.2. 1st Israeli Users Meeting, 26th October 2008 30

4.3. 1st Jordanian Users Meeting, 6th November 2008 30

4.4. 1st Greek Users Meeting, 9th February 2009 31

4.5. 1st Turkish Users Meeting, 20th February 2009 31

4.6. 1st Lebanese User Meeting, June 19th 2009 32

4.7. 1st Syrian User Meeting, July 6th 2009 32

4.8. Life Sciences Thematic User Meeting, November 5th 2009 33

4.9. 2nd Israeli Users Meeting, 1st December 2009 34

4.10. LinkSCEEM International HPC Conference, 6-8 October 2009 36

Discussion 39

5. Assessment of Regional HPC Resources 41

6. SESAME – Assessment of Needs 42

7. Conclusions and Outlook 43

Executive Summary

The Assessment of Needs process has been ongoing throughout the LinkSCEEM project. The two main goals of the Work Package were to assess the needs for HPC in the Eastern Mediterranean and to assess the existing resources and their current utilisation. The major components of this process have been the establishment of a users' database, conducting a number of surveys, holding panel discussions during Users' Meetings and performing an assessment of regional HPC resources.

The current users' database contains some 599 entries and this has been used repeatedly in the context of this Work Package. This database covers a broad range of countries, positions and scientific areas. Major contributions to its size come from the Users Meetings which have facilitated direct access to communities and, in turn, provided them with a means to actively interact with the project.

It has also provided the means to conduct the simplified and detailed resource surveys of the project. These surveys have provided an accurate description of the abilities, expectations and desires of the potential user base of a regional HPC center. This description is, not surprisingly, broad but general requirements are relatively clear. This in turn has influenced the decision process in the selection of the first regional-scale HPC hardware selection of CaSToRC.

The subsequent surveys held after the LinkSCEEM International Conference and the 3 day HPC Workshop in Amman serve to underline an issue that was also a core topic in all panel discussions of the Users Meetings. This was that, clearly, the desire and the need for HPC facilities exist within the region and the research community is hungry for collaboration but hardware facilities cannot succeed without a sustained education and training program that reaches out to existing and future users.

This topic was raised repeatedly during the Panel Discussions along another fundamental consideration, the necessity for the improvement of regional network connectivity. Many researchers feel regional connectivity is crippling a potentially deep well of research capability. This is also critical when considering remote use of European-scale HPC resources. It seems that the consensus among users is that this problem is more political than technical. The solution requires both large investments by governments and cooperation among them.

That such topics are already being addressed within other Work Packages of the LinkSCEEM project is an important validation of the concept of CaSToRC as a regional HPC facility and Center of Excellence for computational science. It is also clear that CaSToRC must actively introduce the regional scientific community to HPC. It should not only provide an educational structure that addresses all levels of experience but also create a user policy and process that encourages usage, particularly for new users.

In the future, the networking process initiated within LinkSCEEM will enable the gradual development of an incubation process that will produce concrete collaborative research and educational thrusts; these will act as magnets and facilitate further networking and engagement of the user communities. The continuous monitoring of user needs will remain necessary, and can be carried out in parallel to this ongoing networking process.

1. Introduction

As foreseen in the Description of Work, the assessment of needs process was based principally on surveys carried out online or through direct requests for information, on the organisation of user meetings and on the setting up of user contact groups. This was originally planned to take place during the early phase of the project, however it soon became apparent that such a process should in fact be continuously ongoing, and in particular that since user meetings and workshops would continue taking place, for purposes notably of outreach, additional input would be generated. It was therefore agreed that the Work Package report would be reworked and updated, using the input obtained over the entire lifetime of the project.

The benefit of this is manifest in the final version of the report in many different ways. The two main goals of the project were to assess the needs for HPC in the Eastern Mediterranean and to assess the existing resources and their current utilisation. Allowing the Work Package to span the entire project has allowed the definition of these two goals to mature. The User contact groups have grown steadily with the number of User Meetings and solidified through the International LinkSCEEM HPC Conference. One can now interpret the resources of the region not just in terms of computational hardware but also in terms of human resources, education, network connectivity and training resources. The panel discussions from the User Meeting have allowed the community to voice their opinions on these matters, and summaries thereof have been included in the present report, as an interesting insight of the user communities' own views on their needs, and the related demands and comments.

The present report reviews the input generated by the various activities conducted within this Work Package:

- the setting up of a user database, and the picture of the regional user community
- the surveys of user needs: in addition to the two surveys conducted in the early phase of the project, two additional ones, focusing primarily on training needs were conducted later
- the user meetings, and notably the panel discussions, that allowed for direct interactions with the regional end-users

2. Users Database

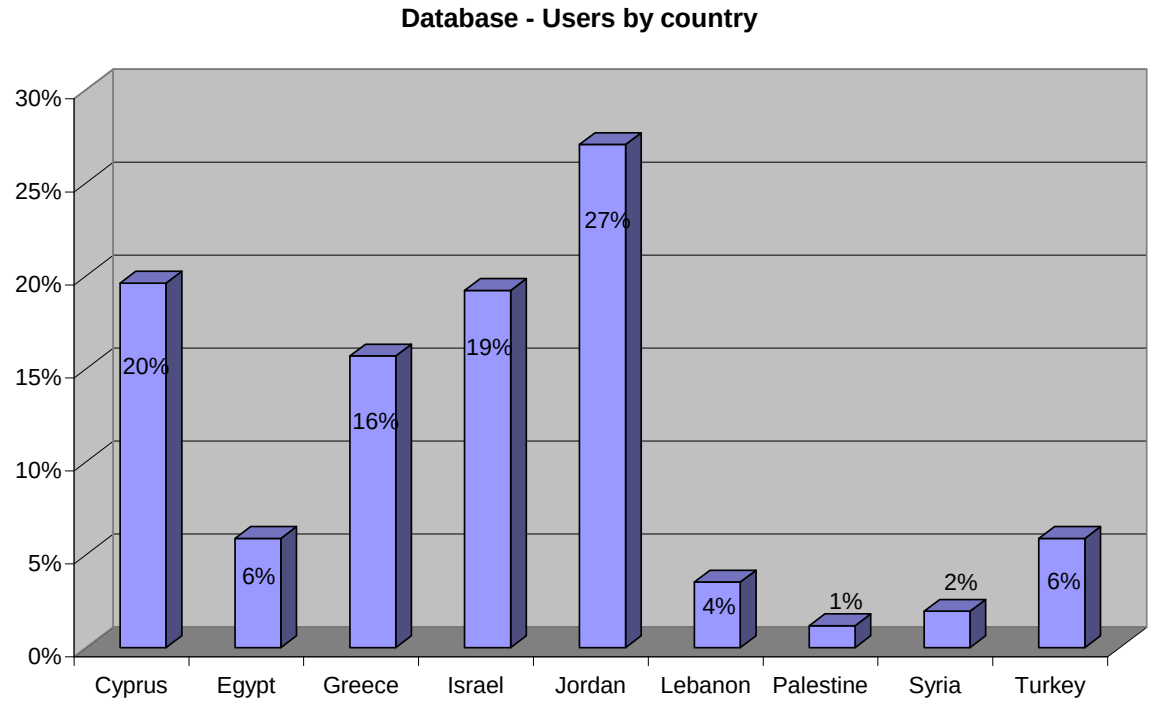
The networking process to potential HPC users in the Eastern Mediterranean has progressed in parallel with the contacts with participants of the users' meetings, it is also of interest in it is own right, as a first step of the build-up of the future CaSToRC user community, and in the short term for the setting up of a user database. The Users database has been continuously updated throughout the project, and used to keep current or potential HPC users informed of the activities of the LinkSCEEM consortium, and of the development of CaSToRC.

Local user contacts groups - i.e. core groups of users interested in being involved in the networking process, in co-organising meetings, channelling information to other potentially relevant contacts, etc. – have been identified in various countries of the region. The advancement of this process has been – unavoidably – quite diverse, depending on the specific context of each country. While the process has been almost immediate in certain countries (Egypt, Israel, Jordan), it turned out that more time was needed in other cases (Lebanon, Turkey). At this stage the project has active and involved contact groups in all countries in the region with the only exception being, perhaps, the Palestinian Territories. The input and/or feedback obtained from these core user groups (e.g. through private discussions) largely overlaps and confirm with those emanating from the panel discussions, presented in section 4.

2.1. Database as of January 2009

As of January 2009, the Users Database contained 347 contacts from 9 countries of the Eastern Mediterranean (Cyprus, Egypt, Greece, Israel, Jordan, Lebanon, Palestine, Syria and Turkey). Its composition at that time provides a preliminary picture of the future CaSToRC users' community; it is worth mentioning a few of its features:

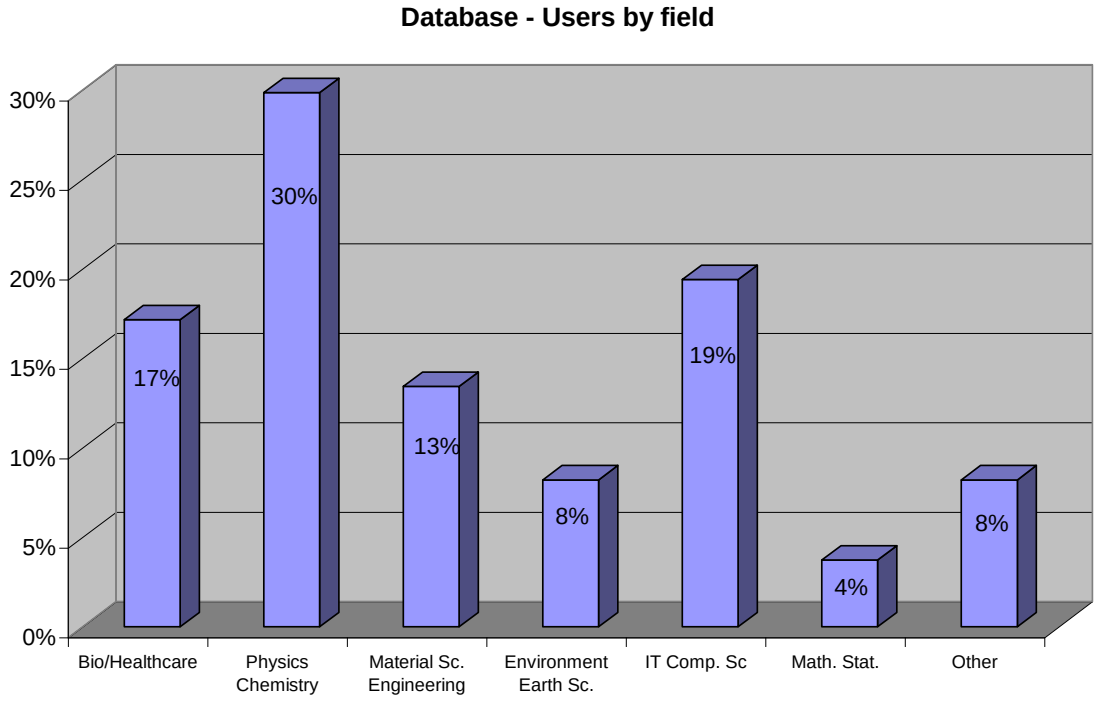
Distribution by country:



The distribution indicates that some interest exists in all countries of the region; it seems to be influenced by the following factors:

- Firstly of course the size of the scientific community, which explains e.g. that the number of Greek or Israeli contacts is, not surprisingly, larger than that of Syrian or Lebanese ones.
- The influence of the users meetings, which greatly facilitates the networking to a larger number of users. This is apparent for instance in the large number of Jordanian contacts, most of which were identified thanks to the users meeting that was held in Amman; conversely, the relatively small number of Egyptian contacts – a country with a large scientific community – is expected to increase significantly when a general users meeting is organised.
- The case of Cyprus is naturally special, since the proximity makes it easier to identify and contact potential users.
- The case of Turkey entails a political specificity that makes it somewhat more delicate than others; it is encouraging however that promising contacts already exist, and that a users meeting has been co-organised with Sabanci University which occurred in February 2009.

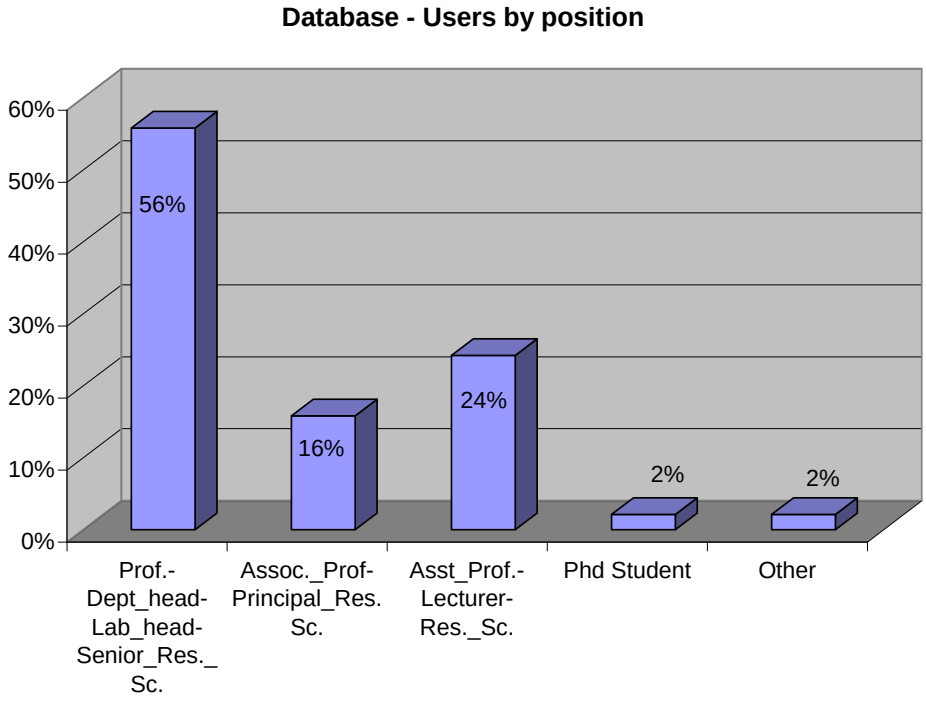
Distribution by field:



This chart shows that the networking process has reached potential users in a great variety of fields; it is worth noting however that a few important areas are not (or only marginally) represented:

- Digital Heritage (i.e. the applications of IT to Cultural Heritage); this should be addressed in coordination with CyI's Science and Technology in Archaeology Research Center (STARC). In particular STARC is the coordinator of the STACHEM project, an FP7 Support Action that fosters scientific networking in the fields of Digital Heritage and Archaeological Sciences; there are obvious synergies with LinkSCEEM, and joint activities have been carried out.
- Economy/Finance/Management; this probably would have required a specific effort, possibly through the organisation of a thematic meeting.
- The private/industrial sector, which would probably require the carrying out of different networking activities, using specific channels. However some of the project's outreach activities (e.g. within the International Conference) should help raise the awareness of decision-makers from the private/industrial sector in respect of HPC. Regardless, the support demand of non-academic activities requires levels of maturity of systems that have not yet been established.

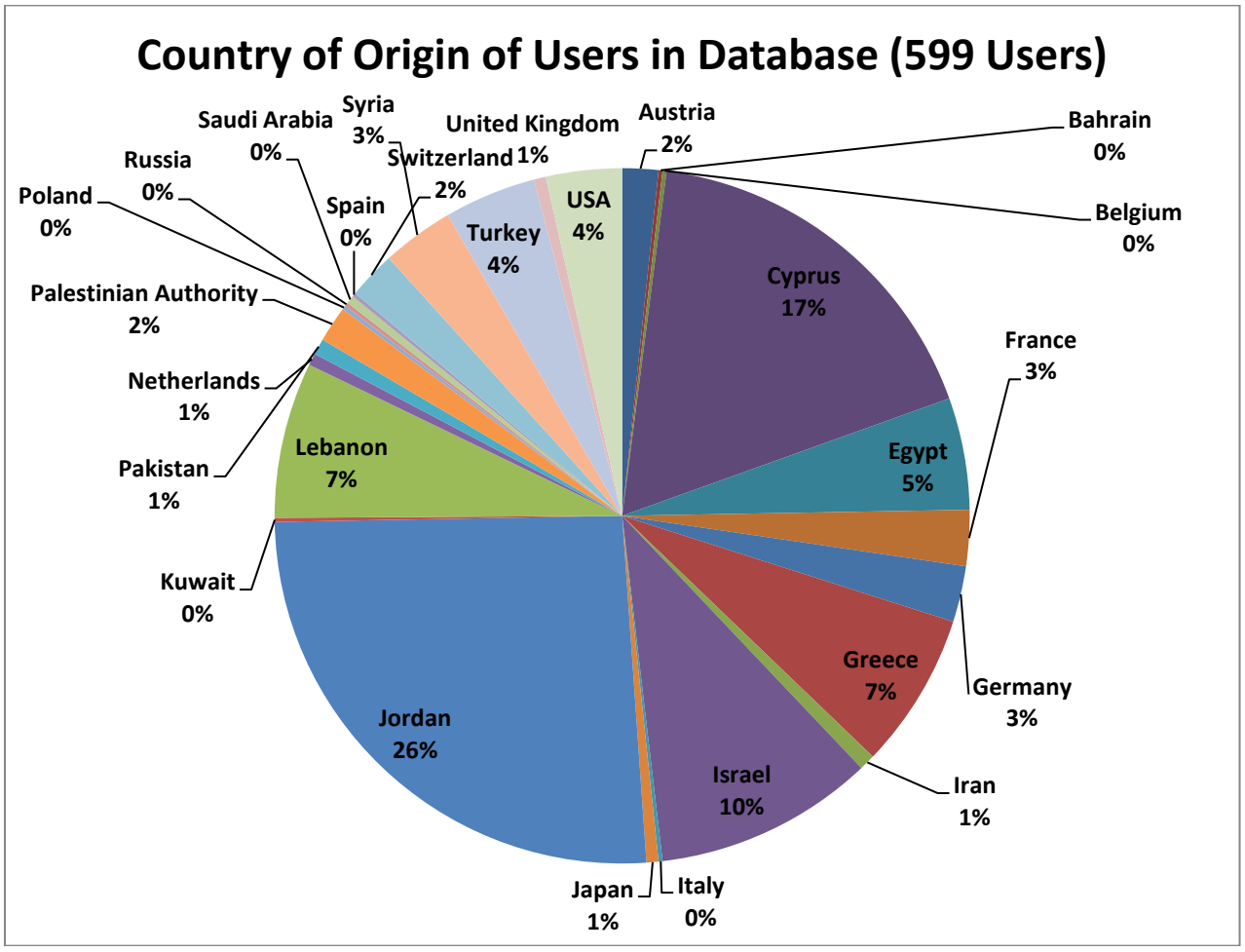
Distribution by position:



This chart shows that all levels of the academic pyramid have been reached, however the predominance of relatively senior scientists seems excessive, and shows that the process is not optimal in this respect. Although it is certainly good to create and maintain strong links with scientific leaders, this chart sends a clear sign that an enhanced effort is needed towards young researchers and students. This should probably be done at grass root level, through various actions that can facilitate direct contacts with young scientists: users meetings and seminars (making sure that the announcements are made visible to young scientists), setting up of training programmes focusing on young researchers (including contributions to existing ones, tutorials that could included within meetings, and online training), and targeted outreach activities.

2.2. Current Database

A database of some 599 contacts has now been created within the lifetime of the project. This has been used to make contact with the User communities with respect to activities relating to project (in particular Users Meetings and the LinkSCEEM Conference) and updates on developments within the project (such as initial findings of the network study and access opportunities to other infrastructures within Europe). It has also been the means to communicate more general information such as Newsletters on the project.



The database is being continuously updated with additional information. In its current status additional information regarding the distribution of the users in terms or subject or position is not representative of the database as this information is not available for a large portion of entries (or categorisation of entries is not currently consistent). Qualitative indications from the database, the conference participation and current activities are that an increasing number of students are being reached. Focal areas are being addressed resulting in thematic User Meetings and these are also now well represented in the database.

3. User surveys

During the first stage of the project, two user surveys have been conducted:

- A simplified survey based on the circulation of a simple form requesting basic information about research interests, current use of computational resources, need for additional resources, and interest for collaboration and/or training in computational sciences.
- A detailed survey, requesting more detailed and technical information regarding HPC applications, to be provided online by participants

Requests for participation in both surveys were made to all the contacts identified within the database at that time, and directly to the participants in the user meetings held in the initial period. As is described below, these two surveys have an interesting complementarity, in that the former has provided limited, basic input from a rather broad panel of potential users, while the latter provided more specific and detailed information about a more restricted panel.

After the holding of the LinkSCEEM International HPC Conference, an additional opportunity for conducting a survey on the participants was exploited. This was a web-based form with much in common with the simplified survey described above and provides a good point of comparison between the outlooks at the initial and latter stages of the project.

In the very final stages of the project a 3 day HPC workshop event was held in the University of Jordan in Amman. Again, a participant survey was conducted designed to gauge the effectiveness of that activity and the desire for further such activities.

3.1. Simplified survey:

Towards the beginning of project the following form was circulated to all contacts in the users database, and to all participants in the users meetings:

If you are an currently or potentially an HPC user, please take a moment to fill in this form, which will provide us with valuable information on your use of and needs for computational resources, and your interest in related collaborative research and training.

Thanks in advance.

Title		**First Name**		**Last Name**	
Position		**Organization**			
Scientific field				**Email**	
Telephone		**City**		**Country**	

Research interests (personal and/or within your institution):

Current use of computational resources (personal and/or within your institution):

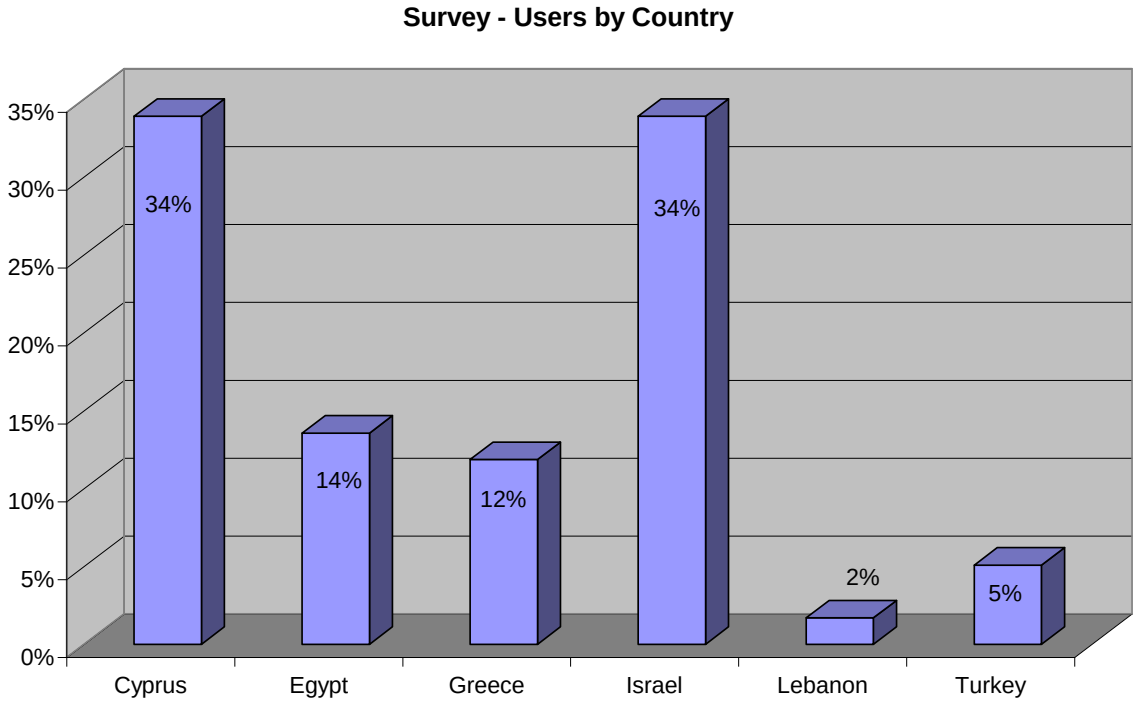
Needs for additional computational resources (personal and/or within your institution):

Interest for collaborative research in computational science (personal and/or within your institution):

Interest for training in computational science (personal and/or within your institution):

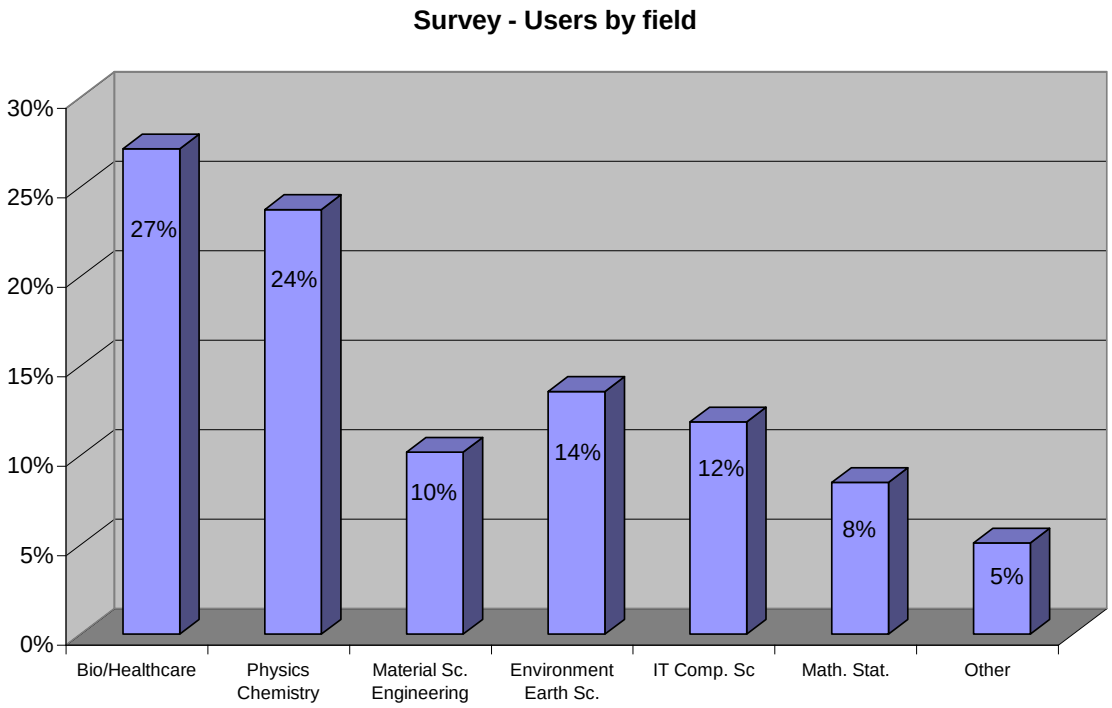
Answers from 59 users have been received; the survey participants form a subset of the user's database, with the following features:

Distribution by country:



Most of the comments that were made above about the database could apply again, with the additional remark that the influence of users meetings seems even stronger for getting feedback from potential users, as can be seen from the comparison between Cypriot and Israeli participation on the one hand, and Egyptian and Greek on the other (where had and had not been held at the time of the survey). However that makes the absence of any participation from Palestine, Syria and especially Jordan, in spite of the users meeting that was held in Amman, even more striking. It seems that the reaction of potential users from these countries to this request for information is somewhat different, and rather cautious; this is also true of other solicitation, as we will see below.

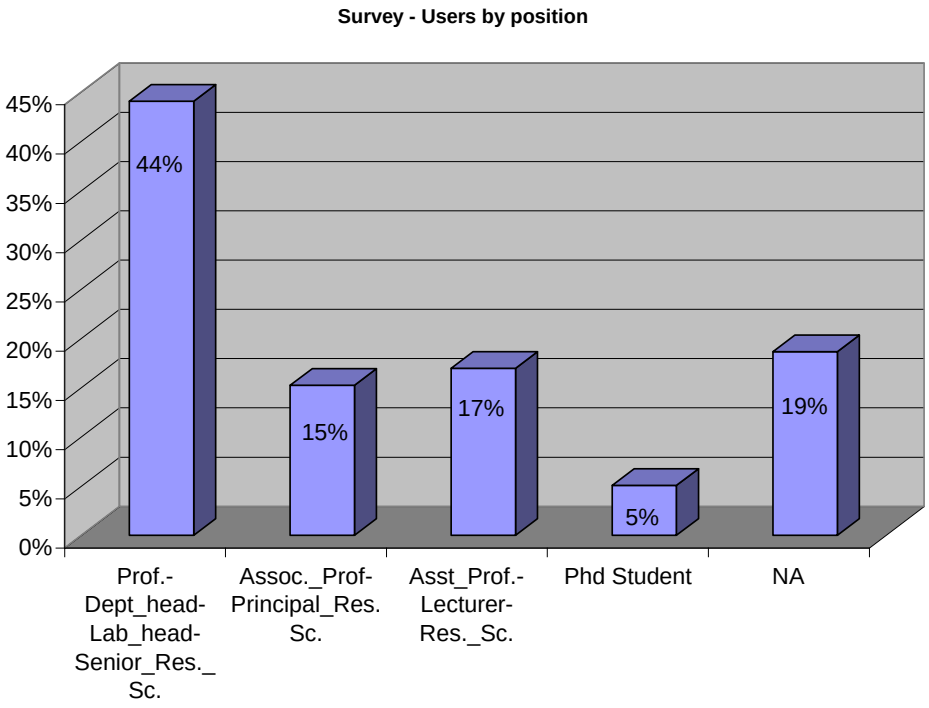
Distribution by field:



The comparison with the corresponding chart in the description of the users database (see above) shows that the fields of Bio/Healthcare and Environmental Sciences have a stronger representation within the survey participants than they did in the database, while the opposite is true of Physics/Chemistry and IT/Comp. Science; one may speculate that this is due to the fact that more scientists in the former two fields, with respect to the latter, have computational needs that cannot be met by their current resources, which constitutes an obvious motivation for participating in the survey.

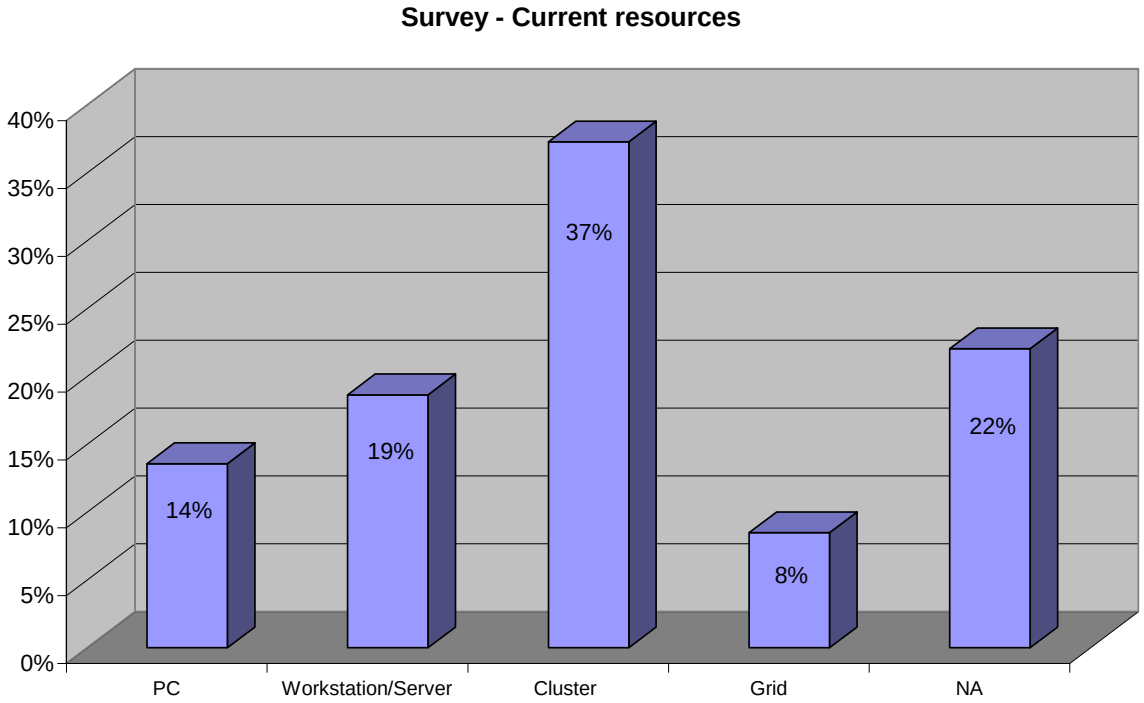
It would be of interest to carry out a cross analysis of the answers to the last three questions in the form within each separate scientific field and/or within each country, however a larger number of participants would be required for such an analysis to be meaningful. Another factor that may be at work here is that, as we will see below, the fields of Physics/Chemistry and IT/Comp have a strong representation within the participants in the detailed survey, which may have led them to neglect the simplified survey; conversely, Bio/Healthcare and Environmental Sciences have a much smaller representation within the detailed survey.

Distribution by position:



This chart is very similar to the corresponding one in the description of the database, with only an enhanced representation of graduate students. This seems to indicate that all levels of the academic pyramid have more or less equivalent willingness to provide feedback to such surveys. There is an increased percentage participation from students which reinforces our impression that a specific effort is in order towards young scientists.

Current resources:

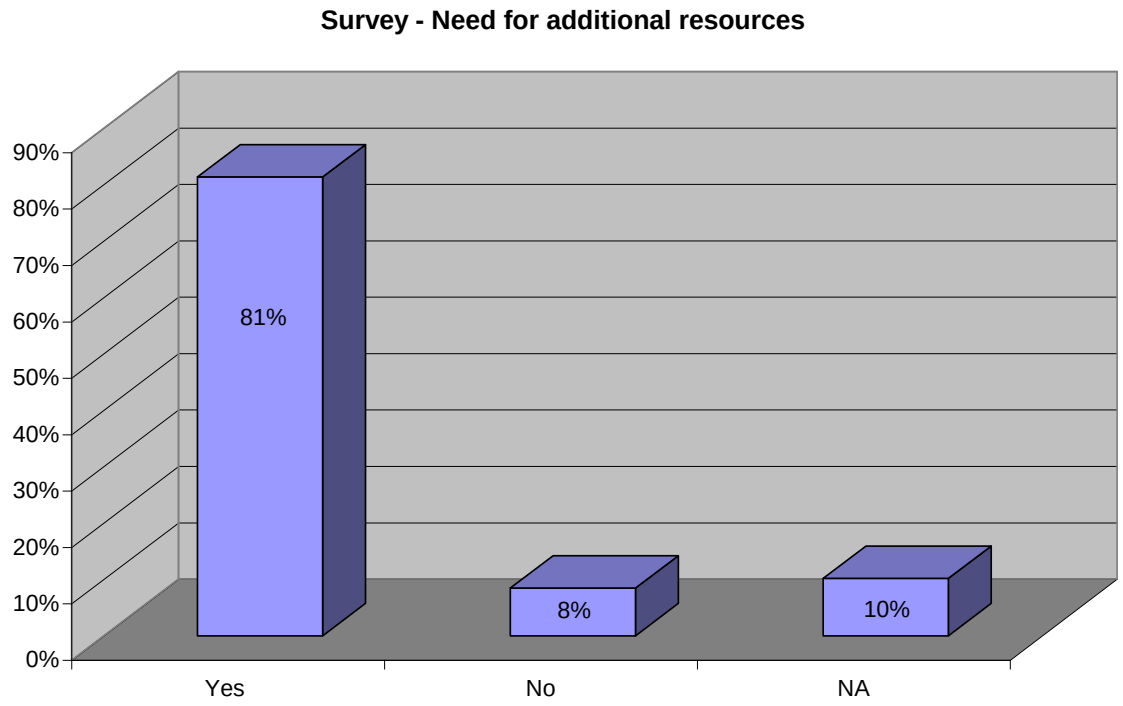


The main indication of this chart is that only 37% of users have access to cluster-type facilities implying that the majority of users only have access to very limited computational resources. This is confirmed by the answers to the next question, which shows that a vast majority of them feel the need for additional resources. The detailed survey gives complementary information in the same direction (see below), as does the conference survey (also below).

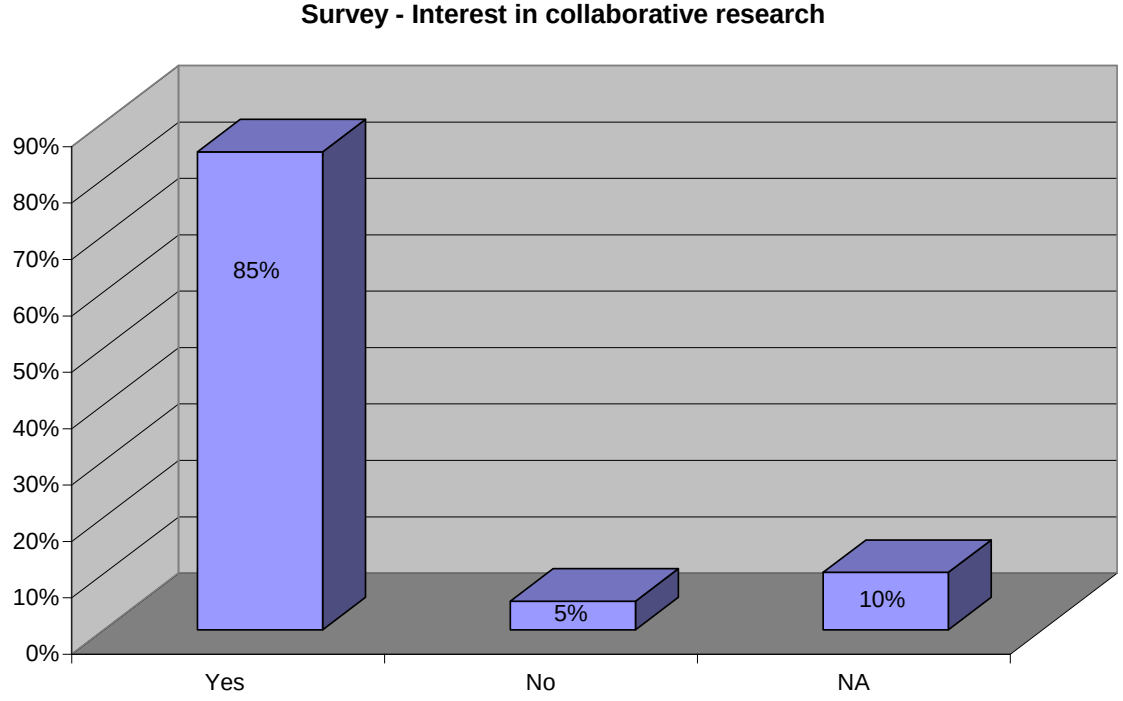
Need for additional resources:

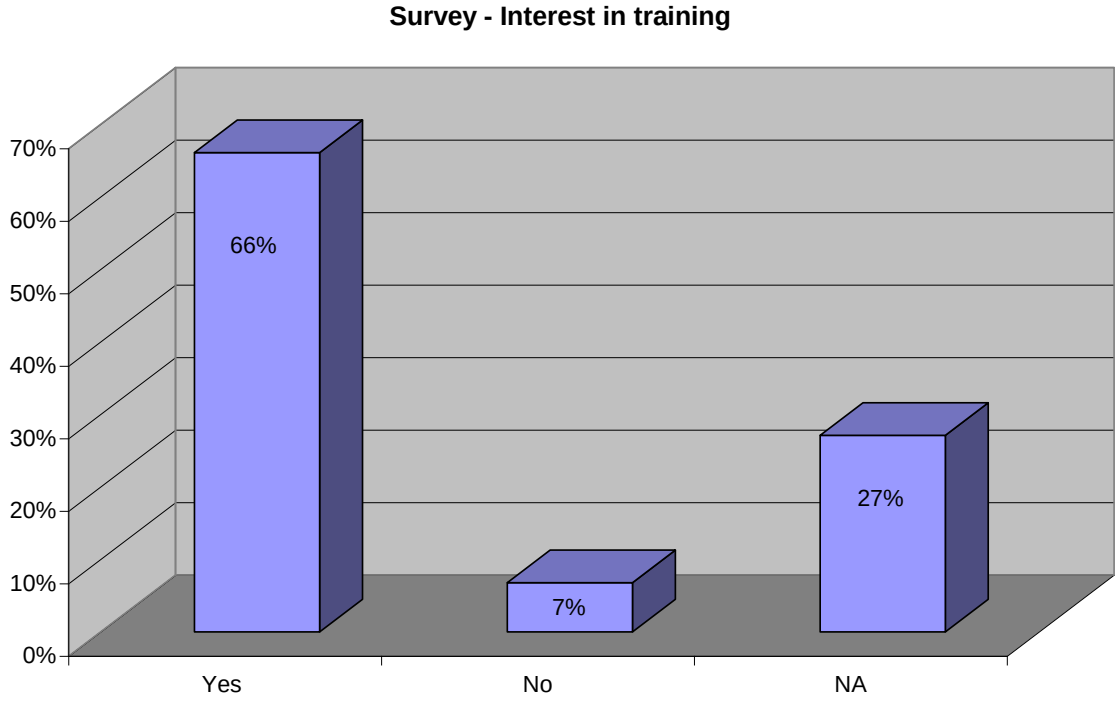
As mentioned above, an overwhelming majority (81%) of survey participants express a need for additional computational resources (see the chart on the next page), which is of course an encouragement towards the development of the CaSToRC infrastructure and the associated networking process, and an important validation of its concept as a regional center of excellence.

Strictly speaking this distribution concerns the participants of the simplified survey, but is very likely true of the users database as a whole, at least to a significant extent. Of course one may object that there is an obvious bias in that extrapolation, in that users who do not feel the need for additional resources may be less interested in participating in the survey. Naturally this initial picture should be confirmed by obtaining feedback from a larger panel of potential users, but it is not expected that this would bring about radical modifications. It would also be of interest to monitor the evolution over time; in particular it is likely that actions directed at potential users to raise their awareness of HPC and the possibilities it offers for their research, together with the enhancement of their expertise through training programmes, will result in a significant increase of their perceived needs for computational resources.



Interest in research collaboration and training:





An overwhelming majority of survey participants (85%) expressed interest in collaborative research in computational sciences, and a somewhat less overwhelming but still significant majority (66%) interest in training. Again that is an important validation of the regional networking process, and of the fact that the CaSToRC infrastructure will be in a position to act as a magnet and catalyst for regional scientific cooperation.
The difference between the interest in collaborative research and training is certainly influenced by the fact that the survey participants are predominantly senior and/or experienced researchers, who are naturally less interested in training; it may well diminish once the networking process succeeds in reaching more young researchers and students.

3.2. Detailed survey

The detailed survey was based on the following questionnaire, implemented online via the Zoomerang website:

1. Name of Application
2. Please provide a brief description of the application code including what types of problems are addressed with this application?
3. Who develops and maintains this code?
4. Is the code publicly available?
5. What organization supports the program?
6. Please describe the general user community(ies) of this application (who, where, number of users,etc.).
7. Give some specific scientific breakthrough(s) enabled by this code (cite publications). These can be either by you or by other members of the user community.
8. What specific scientific objectives/breakthroughs do you have planned while using this code in the next 3-5 years?
9. Do you need to have local access to data sets maintained for persistent access?
10. Approximately how many lines of code does the application currently have?
11. What methods of parallelization do you use?
12. What parallel programming models do you use?
13. What is your parallelism strategy?
14. Does your application include operations suitable for vectorization?
15. Do you have sequential parts of your code that cannot be parallelized?
16. What HPC systems have you run this code on?
17. What is your preferred HPC system and why?
18. For large application runs, what is the typical upper bound on the number of cores on which you run?
19. What is your typical total memory size requirement?
20. What is your typical per core memory requirement?
21. What is your typical run time (wall clock time) for application runs?
22. Characterize your file space requirements below for input, temporary, and output files.
23. Characterize your file sizes below for input, temporary, and output files.
24. For cluster-based applications, what is your node-level I/O bandwidth requirement?
25. What is you global I/O bandwidth requirement for most application runs?

Characteristics of execution time for PLANNED executions. Please provide information on requirements you anticipate while using this application in 3-5 years.

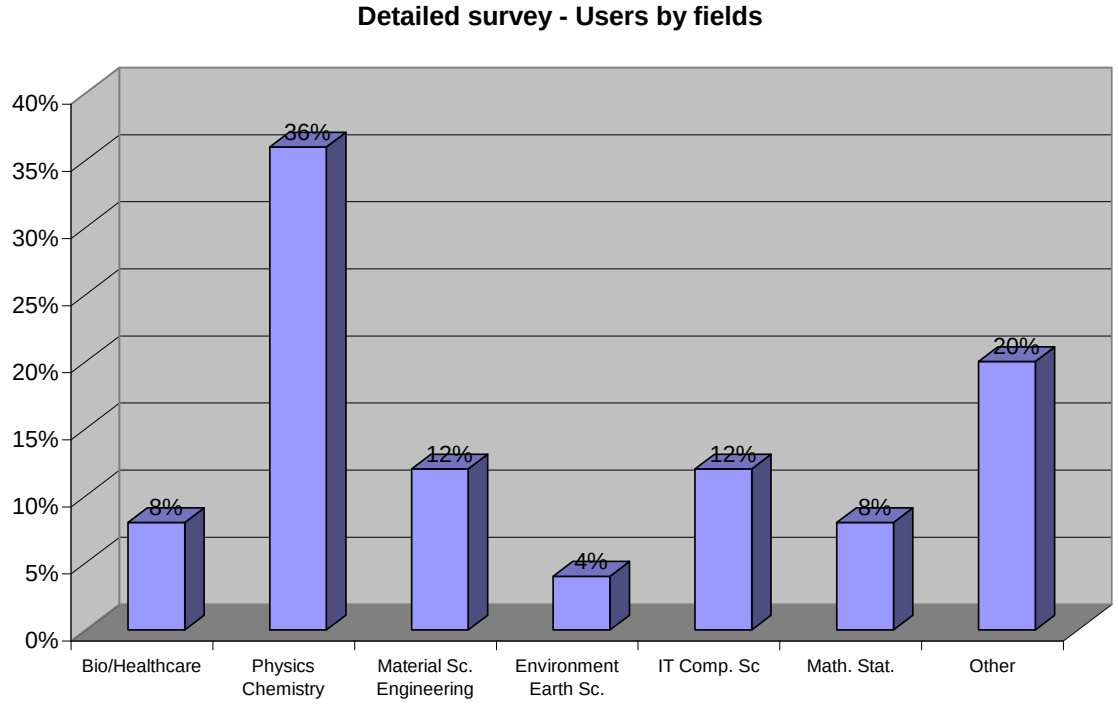
26. What HPC systems would you like to run this code on?
27. For large application runs, what is the expected typical upper bound on the number of cores on which you would like to run?
28. What is your expected typical total memory size requirement for most planned application runs?
29. What is your expected typical per core memory requirement for most planned application runs?
30. What is your expected typical run time (wall clock time) for planned application runs?

31. Characterize your expected file space requirements below for input, temporary, and output files for most planned application runs.
32. Characterize your file sizes below for input, temporary, and output files for most planned application runs.
33. For cluster-based applications, what is your expected node-level I/O bandwidth requirement for most planned application runs?
34. What is your expected global I/O bandwidth requirement for most planned application runs?

Data management: Please characterize some or your data management requirements below.

35. Is the majority of the data you work with generated as output from an application run on large scale resources?
36. Do you obtain data from an instrument, detector or similar data source?
37. Do you obtain or update data from remote data stores (e.g. databases, data archives/repositories, data services)?
38. Do you retrieve from archival storage and reuse or analyze?
39. Do you use data mining techniques?
40. Do you make use of databases (even flat files)?
41. Do you need to have local access to data sets maintained for persistent access?
42. If you need to have local access to data sets, what is the total size of these data sets?

As could be expected, the number of participants in the detailed survey was smaller than in the simplified survey: 25 versus 59. The information they provide have been very useful for the other LinkSCEEM Work Packages, and for the design and planning process of the CaSToRC infrastructure and the associated research and educational thrusts.
In particular, they give insight into the current range of applications employed and developed by prospective CaSToRC facility users. These users have a wide variety of interests: industry, medicine, physics, geology, geography, engineering, mathematics and meteorology (Questions 1 and 6), using programs that have led to a number of significant scientific breakthroughs (Question 7). It is worth mentioning that when regrouped into broader scientific fields the distribution of participants is somewhat different from that of the simplified survey, with a larger proportion of IT/Computer Scientists, and also of physicists and chemists, which was to be expected, since these communities are more versed in IT technicalities (by definition, for the former case) than others.



The survey results unfortunately do not provide explicitly distributions of participants by country or position; indirect indications show that, as could be expected, there is proportionally a larger representation of countries in which computational science and engineering is somewhat more advanced (e.g. Israel), and also a couple of representatives of large international collaborations, notably from the CERN/LHC community.

The main indications provided by the survey can be summarized as follows:

The requirements of these users vary from massive parallel computation to visualization and databasing. As a result of this diversity, the computational requirements of these fields also vary dramatically. At present, some applications are implemented on desktop systems whilst others use the most sophisticated systems available (such as the Blue Gene/P). 80% of respondents use less than 128 cores (Question 18) and 82% say they would be satisfied with 256 cores or less in the future (Question 27).

As regards the memory requirements of the applications 96% currently use less than 256GB of total RAM memory (Question 19) and 95% would be happy with this figure in the future (Question 28). Given a 256-node system this is just half the standard per core allowance in modern systems. However, 25% of respondents say they need over 2GB of per node memory (Question 20) currently. This drops to 5% in the future (Question 29) indicating that users are either introducing or upscaling the level of parallelization in their codes.

The question of parallelization is an interesting one and is addressed both explicitly and implicitly repeatedly throughout this questionnaire. Implicitly, it is addressed in Question 3, which asks who develops and maintains the code for the applications. Few have indicated that they write all software themselves, most users (57% or more on interpretation of the responses) use commercial, community or collaboration maintained software. If it is being

actively used and maintained such software is very likely to have parallel capabilities now or in the future.
However, only 44% of this software is publicly available (Question 4). The other 56% must then be made up of commercial and private software. Since only 4 respondents listed commercial software then these must be collaboration based codes. Of these codes, 52% have less than 10 000 lines (Question 10) and a further 19% with under 50 000 lines. In these cases implementation of parallelization should be reasonably quick with basic appropriate HPC skillsets. 33% of users use the Master/Worker parallel programming model (Question 12) and 68% use domain decomposition as a parallelism strategy (Question 13). 60% of users use MPI as the mode of parallelization (Question 11) but there is little indication of the use of OpenMP or hybrid methods. Given the current popularity of multi-core, tightly coupled systems, all of these factors would indicate significant room for training in this particular programming area.

Indeed 81% of users presently use loosely and tightly coupled Linux clusters (Question 16) with few using non-commodity machines such as the Blue Gene. This 81% persists when discussing future requirements (Question 26). Responses to Question 17, when discussing preferred HPC systems, also support this viewpoint. However, 84% of respondents have indicated that sections of their code are suitable for vectorization (Question 14). Such sections are the basis for the current renaissance in accelerator technologies and usage. Here, again, it would seem that there is significant scope for education. Not unexpectedly, 83% also responded that sections of their codes cannot be parallelized (Question 15), the significance of this depends on the time of execution of these sections. 65% of current utilization is under 24 hours of runtime (Question 21) and this is expected to increase to 76% in the future with a further 21% having a runtime of between 1 and 5 days (Question 30). Again such runtimes are extremely compatible with typical administration policies on Linux clusters.

Other important issues are the questions of data storage and network connectivity. Questions on current and future storage requirements again reveal a wide variation between users (Questions 21 and 31). 95 % currently have less than 1TB of input files but 20% of these expect to use more input data in the future. 84 % use less than 500GB of temporary storage and this is not expected to change. When it comes to archival storage, 50% currently use less than 100GB but the other 50% spreads across a very large range with 3 users using between 5TB and 25TB and in the future this picture changes little. This data is spread mostly between 10s and 100s of files from medium (100MB to 1GB) to very large files (>1TB) files (Question 23) and the only significant expected change in the future is that some users expect to increase to 1000s of files in the medium size range (Question 32).

The bandwidth requirements for these datasets is currently satisfied by 10MB/s per node in 82% of cases (Question 24) but this drops to 66% when considering future projects (Question 33). 94% of current requirements are serviced by a 10GB/s global I/O bandwidth at present (Question 25) and this remains unchanged in the future (Question 34).

Notably, 86% obtain their data from the output of work on large scale resources (Question 35), only 30% obtain data from an instrument, detector or similar device (Question 36) and 52% obtain or update data from remote stores (Question 38). 73% do retrieve and reuse/analyze data from archival storage. However 61% do not use databases (Question 40)

and 59% do not use data mining techniques (Question 39). About half of users need local access to datasets for persistent use (Questions 9 and 41 which were accidentally created identical) with the size of these datasets being less than 1TB for all but one user (Question 42). It would then appear that very large scale storage resources are not immediately required but good network connectivity is necessary. Again opportunities for education in current data technologies present themselves.

Significant work is underway in this user community in a wide variety of scientific areas that have an impact on topics as varied as nozzle-clogging in the steel industry to quantum chromo-dynamics and insect conservation in Egypt to brachytherapy treatment planning (Question 8). All indications in this survey would suggest that a moderately-sized tightly-coupled Linux cluster with perhaps some accelerator technologies and a medium sized storage system could cater to most users who took this survey. Importantly however, a number of areas where there is a lack of expertise have also been exposed. The users in this group vary from the very small scale to the very large scale with the majority closer to the small scale. This, coupled with the actual number of responses received, would indicate that there is significant room for improvement with respect to scientific computing. Were a program of education in place, it is highly likely that the computational requirements would increase dramatically.

3.3. Conference Survey

The questionnaire below was prepared online and those wishing to access the conference proceedings were required to complete it (the proceedings were also delivered on CD by post). There were a total of 47 responses to the survey. Of these 43 indicated that they would (or might) use the facilities to be provided by CaSToRC. It is these 43 responses that form the basis for the statistics given, since they are a representative sample of the (potential) CaSToRC user community.

Participant Questionnaire

In which country are you based?

Where is your institution?

What is your research area?

Are you a

How would you rate the programme?

How would you rate the location?

Do you see a potential for collaboration with other participants at the conference?

Would you like to make use of the computational facilities provided at CaSToRC (Cyprus Institute)?

Do you feel you currently have the skills necessary to use the types of HPC facilities discussed during the conference?

What is the highest level of computing you currently use at your Institution?

Tutorial Section

Tutorial area attended:

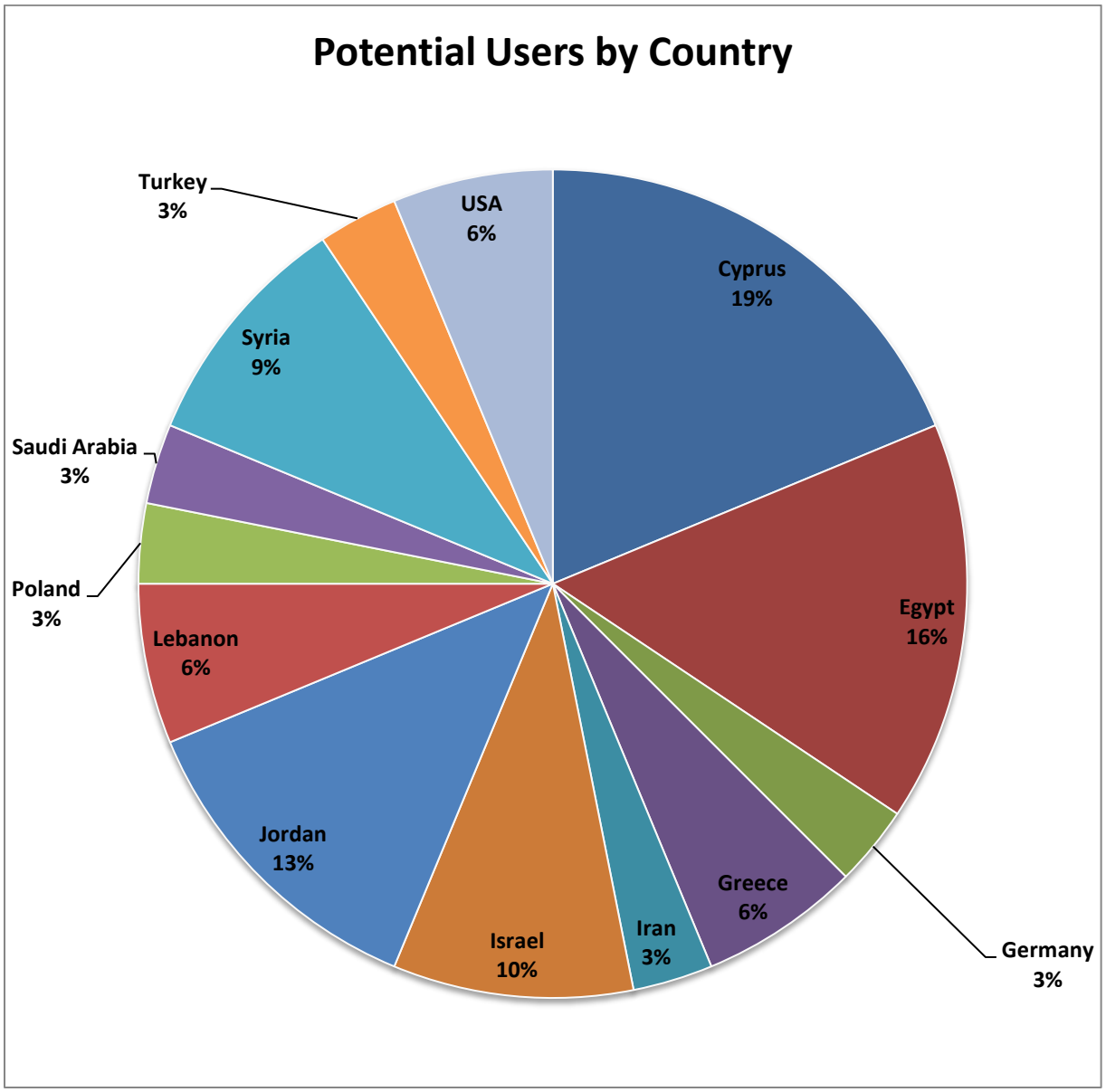
Do you have previous experience in this tutorial area?

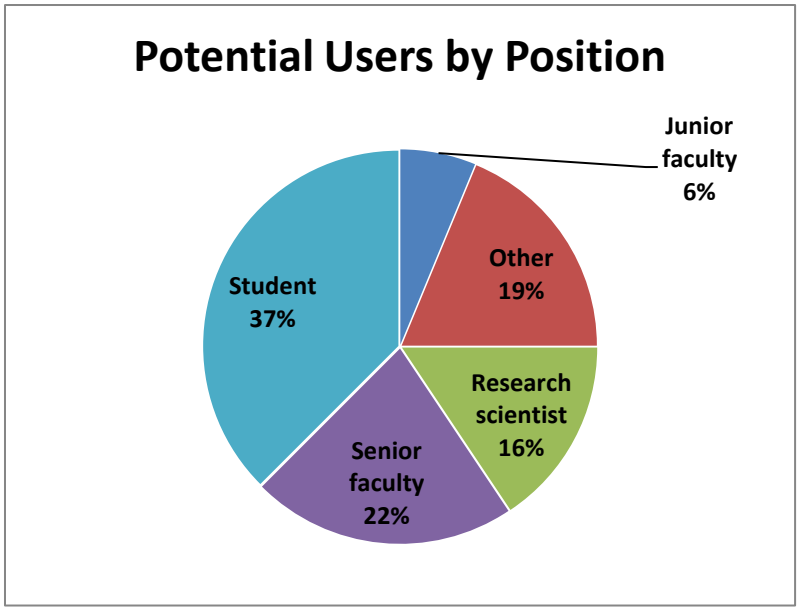
Did you find the tutorial informative?

Potential Users by Country:

The distribution of responses by country holds a reasonable resemblance to the distribution seen in the Users database. Exceptions may be the relative proportion of Jordanians (lower representation), Egyptians (higher representation) and Syrians (higher representation).

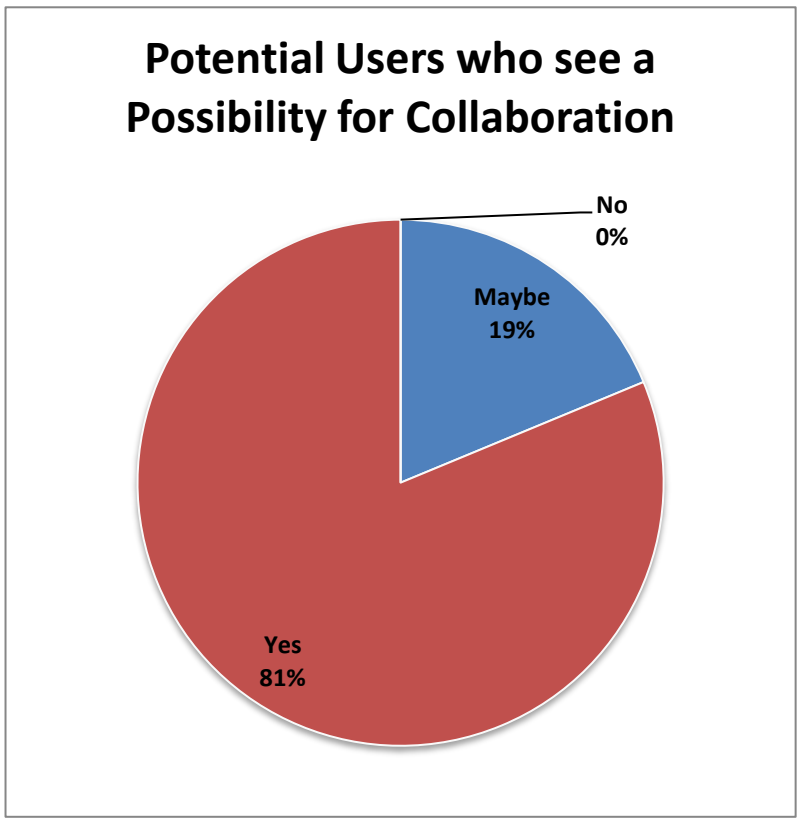
There are also a number of additional responses from other countries in Europe and the US. These are likely to be either people of Eastern Mediterranean origin working abroad or attendees who foresaw prospects for collaboration.





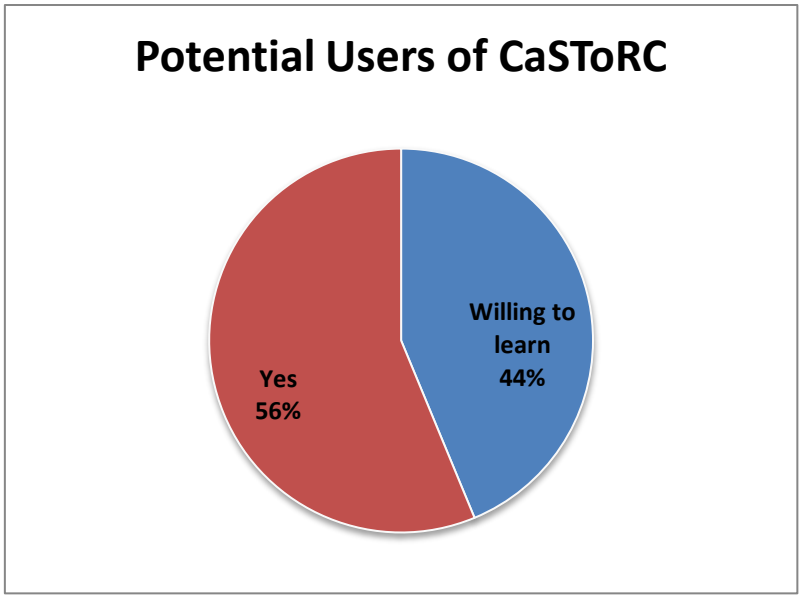
Potential Users by Position:

The distribution by position compares very favourably with the distribution seen in the Simplified Survey. In particular, a far larger portion of students is sampled. This is, at least in part, by design since students were targeted (and in many cases financially supported) for the conference. We also see a good distribution between researchers, junior and senior faculty.

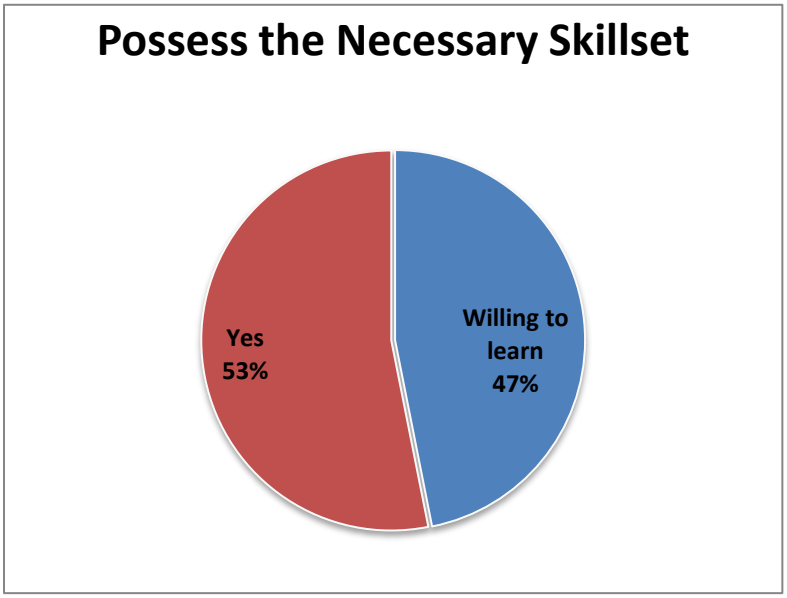


Potential Users who see a Possibility for Collaboration:
It is extremely encouraging to see that an overwhelming majority of conference participants see a definite possibility for collaboration with other participants of the conference. This is a testament to the effectiveness of the LinkSCEEM project as a networking and community building initiative.

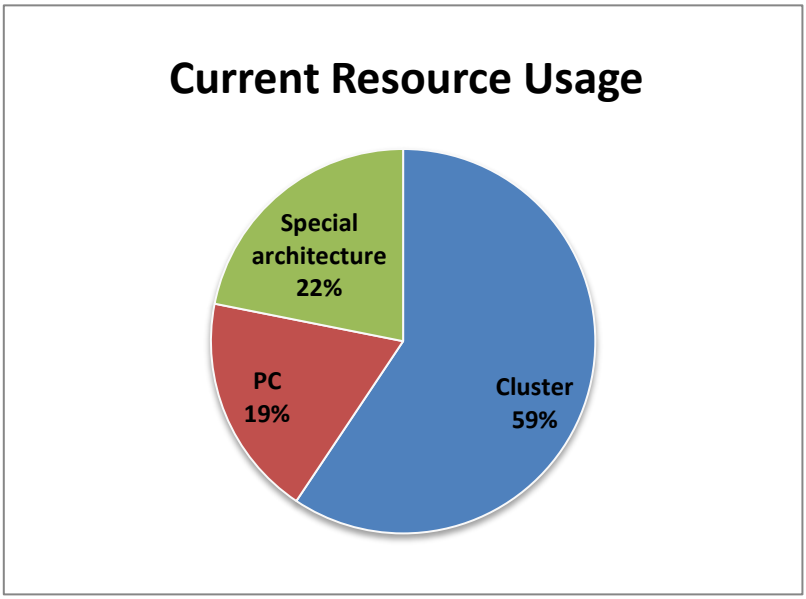
Users of CaSToRC Facilities:
The overall sample was collected based on the fact that the contributions were by those who saw scope for usage of CaSToRC facilities. It is interesting to note however that the distribution in this respect is reasonably balanced between those who feel they can do so immediately (56%) and those who feel they must go through a learning curve to do so effectively (44%).



This is again evidenced by the response to the subsequent question regarding whether the participant felt they had the skillset to effectively utilise modern HPC facilities. The distribution is almost identical to above and highlights that CaSToRC must effectively address its role as an educator in order to effectively engage its potential user base.

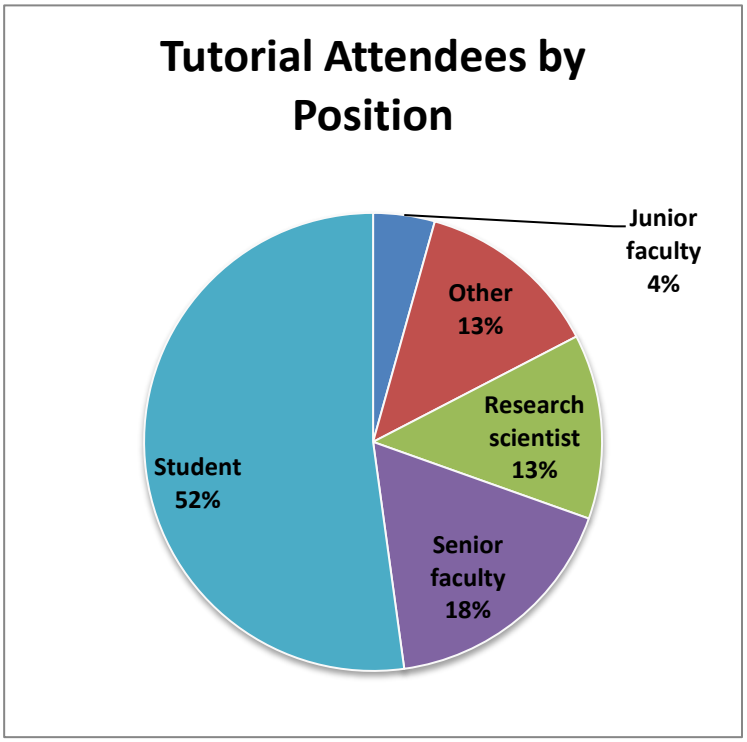


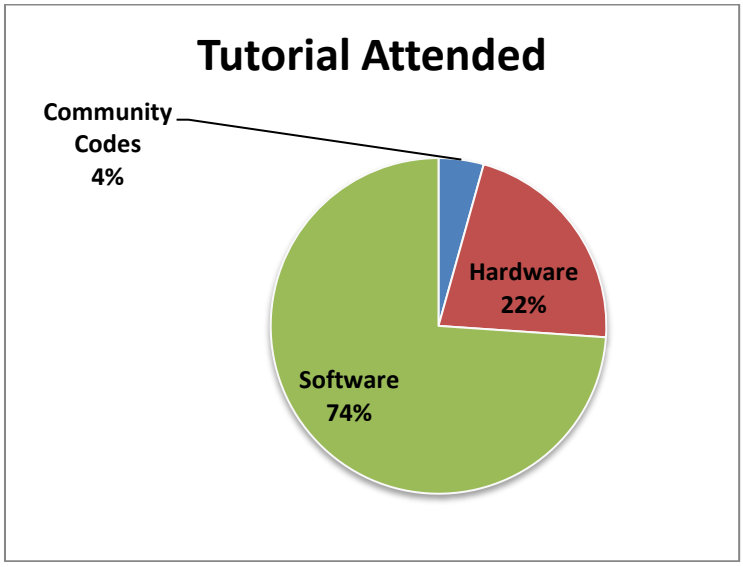
Current Resource Usage:
There is an interesting overlap between the above responses and the question as to what the current resource usage level of the participant is. While 19% still only utilise Personal Computers, 59% are familiar with cluster systems and 22% with special architectures (such as Blue Gene, GPU, etc.). This would indicate that people's perception of the facilities at CaSToRC would go beyond that of a standard cluster system (given that only 19% would not have adequate skills for that type of system and yet 47% feel they require training).

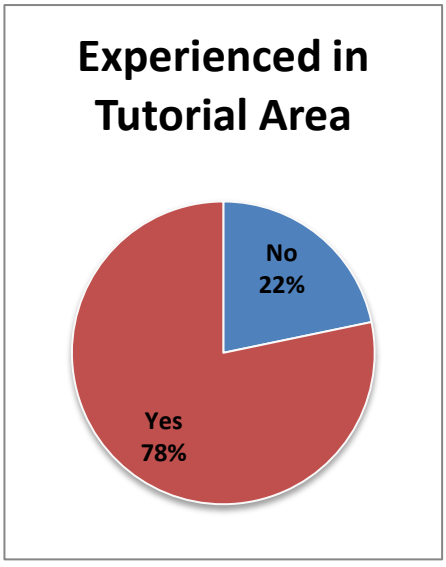


Responses from those who attended tutorials:

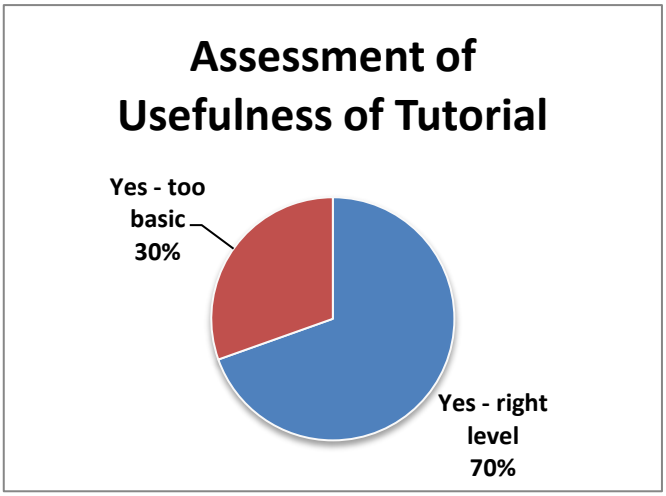
Of the 43 responses used in the overall sample, 23 of these respondents also attended a tutorial session. These responses were almost exclusively from the region and we can argue that they again have a very favourable distribution between students and more senior researchers and faculty.
Of these, 74% took the software tutorial, 22% the hardware tutorial and only 4% the community codes tutorial. It should be noted however that the community codes tutorial only addressed one subject area and was therefore unlikely to appeal to many of the attendees from other areas.







Of these, 78% had previous experience in the tutorial area. The tutorials themselves were aimed at novices and this is perhaps why 30% responded that they found the tutorials useful but at too basic a level.



3.4. 3 Day HPC Workshop Survey

Between the 19th and the 21st of January 2010, CaSToRC conducted a LinkSCEEM 3 Day HPC Workshop, hosted by the University of Jordan. The workshop was given by 3 members of CaSToRC staff: Patrick Fitzhenry, Alan O'Cais and Alexei Strelchenko. It was seized as an opportunity to gauge the capabilities of the User Community and assess the level of material necessary to engage and educate that community. It also provided another opportunity to engage the community on a practical level and make them aware of the assistance that exists for them.

This was also an opportunity to obtain feedback on how the workshop catered to participants needs and what they thought those needs may be in the future.

Evaluation Questionnaire

1. Are you a: (Please circle)
   - Student
   - Researcher
   - Other (Please Specify)

2. Did you find the course interesting? (Please circle)
   - Very Interesting
   - Interesting
   - Average
   - Not interesting
   - Boring

3. What did you find most interesting?
   - MPI
   - GPU
   - Using a cluster
   - Programming
   - Other

4. Did you find the level of the material?
   - Too hard
   - Hard
   - Average
   - Easy
   - Very easy

5. Do you think what you have learned will help you in your research/work?
   - Definitely
   - Yes
   - Maybe
   - No
   - Don't know

6. Would you like to attend further workshops or similar activities?
   - Yes
   - No

7. Do you have any suggestions to improve workshops like this (e.g., more/less lectures, more/less tutorials, other material, community codes)?

There were a total of 16 responses received from participants from universities in Amman: 8 from researchers, 6 from students and 2 who described themselves as 'other'. 14 of these found the material both interesting or very interesting, with one finding it average and another not interesting. 8 found the GPU section most interesting, followed by 5 who found MPI most interesting. Significantly, all of the responders would like to see more workshops conducted in these areas.

Perhaps the most the most telling portions of these questionnaire were the suggestions for their future improvement. To highlight some of the issues raised some of those comments are included here:

- "Unix administrators ... should attend", "We need another workshop about making an MPI cluster", "prefer to have more cluster administrative sessions" – this highlights that some feel that the infrastructure of their universities is not prepared to support HPC requirements. They are seeking assistance in setting up and administering these environments and this is clearly an area where CaSToRC can assist.
- "Community codes will help in understanding the whole subject better" – it was highlighted to attendees that there can easily be community software that requires only slight modification to meet their needs and they have clearly responded to this.
- "We need more workshops and training" – many attendees repeatedly asked for additional training and tutorials, both in person and online.

4. Panel Discussions of User Meetings

The objectives of the user meetings were:

- To provide a forum where users and providers of high-performance computing (HPC) facilities in the Eastern Mediterranean can meet and exchange views and information;
- To explore the available HPC infrastructure in the region, and the ways it currently serves the user communities;
- To provide an opportunity for computational scientists from the region (both users and providers) to learn from the experiences of leaders of HPC centers in Western Europe and the US;
- To lay the foundations of a user community around the planned HPC facility of the Cyprus Institute's Computation-based Research and Technology Research Center (CaSToRC).

The format of these meetings was based on the following model: a first session including a few presentations by LinkSCEEM partners (e.g. one on the project itself, one on the regional HPC scene and the development of CaSToRC, and one or two other(s) on other relevant subject(s), such as "The user community and applications portfolio of a tier-1 HPC center"), followed by a second session including a number of short presentations by users in various scientific fields, describing their research interests and projects, and the related computational needs. In addition it is useful to ask one or several leading figure(s) of the local scientific community to present an overview of the local computational science & engineering scene.

Other meeting formats, such as thematic and by scientific field, where the LinkSCEEM User Meetings form part of a larger program, have also been utilised to further the goals of project and expand the user base by addressing whole communities.

Significantly, the Users meetings have provided a means of direct access to potential users of CaSToRC facilities. Of particular relevance to the LinkSCEEM project are the round-table and panel discussion sessions which have been held after the presentation sessions. Repeatedly, these sessions held during User Meetings have yielded information relevant to this work package. For this reason, while the meetings themselves are essentially a dissemination and outreach activity and belong to WP7, the report on panel discussion sessions of the User Meetings are reported here due to their direct relevance to this Work Package.

4.1. First Users Meeting of the LinkSCEEM Project, 8th April 2008

Venue: Guy Ourisson Building, CyI Athalassa Campus, Nicosia Cyprus

Theme: *The future of scientific computing in the Eastern Mediterranean*
Co-chairs: Prof. C. Alexandrou (IGB Chair, CaSToRC) and Prof. T.Dunning (NCSA)

After a welcome to the participants from C.Alexandrou, T.Dunning made an introductory statement explaining that in the framework of the partnership that has been formed between the Cyprus Institute (CyI) and the National Center for Supercomputing Applications (NCSA), it is of great importance that the design and planning effort of the Computation-based Research and Technology Research Center (CaSToRC) be based on an assessment of the needs and demands of potential users in Cyprus and the Eastern Mediterranean region. He argued that it is therefore crucial that potential CaSToRC users provide input on the type of problems they address and applications they use, notably through meetings such as the present one and through participation in various surveys and requests for information. The networking process to potential users should also involve the gathering of information on the kind of users support and training/educational programmes that would be needed, as well as information on the environment of research activities.

The discussion then turned to the issue of the users needs in various scientific areas. It was agreed that in general there is a clear need for more computational resources, as well as for enhanced connectivity and archiving capacities. Computational chemistry and computational biophysics were mentioned by S.Skourtis (University of Cyprus) as a research area of obvious relevance, in which the creation of a research focus group – a possibility mentioned by T.Dunning – would be of particular interest, because of the interdisciplinarity that is needed, and of the variety of computational codes that need to be used for different objectives. C.Alexandrou mentioned climate modelling, which was also agreed to be an obvious focus, with enormous societal importance in the region. In this field very significant input is expected from IUCC/Tel Aviv University, as a LinkSCEEM partner, and naturally also from the participants in climate modelling activities planned at CyI/EEWRC.

The possibilities of structuring the users communities around thematic research focus groups, or around local contact groups in countries of the region were both mentioned. In addition to a continuous dialog between the users communities, the team in charge of the CaSToRC design and planning and the LinkSCEEM consortium, direct interaction and exchanges between users should naturally also be encouraged, as stressed by T.Dunning and A.Stupp (IUCC & Tel Aviv University), notably through the setting-up on the CaSToRC website of an online exchange system (e.g. through a Wiki) where users can add information, exchange ideas, make comments, etc.

As argued by N.Nassif (American University in Beirut), it was agreed that this initial meeting needs some adequate follow-up for a sustainable activation of the users communities. As planned in the LinkSCEEM proposal, other users meeting will be organised on countries of the region, and thematic workshops and summer schools were also mentioned as interesting possibilities. It was argued that in order to succeed the project should be based on a broad scientific community (i.e. not restricted to local users), and make plans on a timescale much larger than the lifetime of the first investment, so as to achieve sustainability.

It was agreed that, as suggested by A.Sharaf Eldin (Helwan University, Egypt), an assessment of current computational resources would also be useful, as well as a review of existing Masters and PhD programmes in computational science and engineering.

The issue of the access of users to the future CaSToRC resources was also mentioned, including the peer review mechanism that will be set up. It was argued that the small scale needs that are encountered in the initial phase of many projects should not be neglected, since serving them adequately can be a significant factor of progress; T.Dunning explained that in many US HPC infrastructure, a specific access channel is available for very small needs (up to a certain threshold), which involves no peer review and a very simple application (one paragraph).

The issue of regional network connectivity is clearly of great importance; the achievements of the Eumedconnect collaboration were recalled by F.Karayannis (GRNET, Greece), and their limitations in terms of capacity and topology were recognised (links only from Eastern-Mediterranean countries to European countries, and not within the region). It was argued that the presence of Jordanian and Syrian participants in the LinkSCEEM consortium could be a plus for finding solutions to the political difficulties that have weighted on this issue.
T.Dunning mentioned that his experience in the US is that in some cases the influence of high level political decision makers (such as State Governors) is needed to make progress on connectivity. Following a remark by M.Gharaibeh (SESAME, Jordan) the creation of the synchrotron light-source at SESAME was agreed to be a regional example of a similar nature, while the example of the conference organised by the PRACE consortium for high-level decision-makers from European industrial firms was mentioned by A.Osseyran (SARA, Netherlands) as an interesting source of inspiration, in particular for the political component of the International Conference that is projected within the LinkSCEEM project.

Following remarks by J.Towns (NCSA), it was therefore agreed that the scientific case for the need for enhanced connectivity should be prepared (climate modelling, synchrotron applications – in collaboration with SESAME - and remote training/virtual graduate schools were mentioned as obvious items). The NRENs and ISPs should be consulted about the related technical requirements, and the case for the upgrade of regional connectivity should then be taken to political decision makers in the region, through the organisation of meetings, of the International Conference and through other outreach actions.
Y.Torman (JUNET, Jordan) argued that political support could reasonably be expected in most countries of the region, but that the decision-making process could be tedious and slow in some cases.

Concerning the training/education programmes that should be set up, it was agreed that they constitute an important ingredient for the build-up of a sustainable users community, in parallel to the users support. The importance of the inclusion in the said programmes of some fundamental aspects of computational mathematics - in addition to the more obvious applications oriented computational science - was stressed by E. Kontoghiorghes (University of Cyprus). Indeed it was agreed that, as mentioned by F.Omara (University of Cairo) it is becoming more and more essential that computational science research get opportunities to interact with computer science and applied mathematics.
The importance of creating synergies with existing regional initiatives was emphasized; this naturally concerns higher-education programmes in computational science and engineering.

A regional collaboration, supported by NASA and ESA, concerning astrophysics data analysis, and aimed at junior researchers, was also mentioned by A.Ibrahim (Cairo Univ. & American University in Cairo).It was agreed that engaging industrial partners in the region will also be important.

To conclude, T.Dunning stated that so many difficult problems need to be addressed for the advancement of science and the welfare of the planet, that clearly all available talents will be needed. It was agreed that precisely the Eastern Mediterranean region can contribute to this common effort its human resources, with partly unexploited reserves of talent, and large numbers of young scientists eager for opportunities to engage in, amongst other things, computation based science and technology.

4.2. 1st Israeli Users Meeting, 26th October 2008

Venue:Tel Aviv University, Tel Aviv, Israel
Co-organisers: IUCC

The main outcome of the meeting was that the infrastructure that is planned by the Cyprus Institute will be beneficial for all the participants. However, it is of major importance for the computation center to be well communicated in order for the resources to be efficiently exploited.

Prof. Moshe answered to a question of Prof. Weill about the reasons that lead to the failure of the attempt to build a central HPC facility in Israel in the past. Prof. Gottlieb answered that the main reason was that the government, although it promised, it didn't fund the project. This raised a discussion on how Israel managed to get good connectivity to the Europe. As Prof. Gottlieb explained, this was achieved through pressure by the scientific community.

4.3. 1st Jordanian Users Meeting, 6th November 2008

Venue: University of Jordan, Amman, Jordan
Co-organisers: JUNET

Network connectivity (both of itself and indirectly through discussions on large data transfers) was a recurring topic throughout the day. Also discussed was the necessity for sufficient stimulation of the user base to ensure full utilization of large scale facilities. It was also argued that systems, both locally and at the Cyprus Institute, should be put in place to deal with capacity as well as capability computing needs since there were institutions in the region where these needs were not being met.

The panel discussion, which consisted of Prof. Victor Jongeneel of the Cyprus Institute, Dr. Norbert Attig of the Julich Supercomputing Center, Dr. Assad Sakhel of Balqa Applied University and Dr. Mohamad Gharaibeh of SESAME, took place after the user presentations. During this discussion Prof. Nabil Nassif reiterated his belief that, based on their experience at the American University of Beirut, significant resources should be placed in the adequate preparation and training of a suitable user base for any proposed facility. Also discussed was the date the SESAME facility would begin operations and what its data requirements would be at that time. There was also discussion on GRID facilities in the area and the possibility of using cycle-scavenging systems to deal with some of the capability computing requirements

at the university level. Prof. Jongeneel commented that multiple solutions for such a system existed but suited only a small subset of applications. Dr. Attig made the point that a smaller scale facility should exist to deal with capacity computing and also as a proving ground for users prior to applying for resources to use the larger facilities.

4.4. 1st Greek Users Meeting, 9th February 2009

Venue: OTE Academy, Athens, Greece
Co-organisers: GRNET

The panel discussion, which ultimately was organised as a forum discussion because of the numbers present, took place after the user presentations. During this discussion a number of important issues were raised. Prof. Halatsis asked about the connection of the Cyprus Institute to the Cypriot and regional economy. Prof. Kalisperis of the CyI explained that an explicit connection would be made in a planned finance-related Research Center. As it stands the research centers have implicit economic connections through the subject matters of their research, such as the environment.

There were a number of questions relating to building design for a Tier 2 computing facility. Prof. Kalisperis stressed that, as well working with companies such as IBM, there is also a proposal currently under preparation which aims to evaluate the optimal data center design specific to the Eastern Mediterranean region.

Two particularly significant issues relating to Cyprus were also raised. The first was the question of how many students the Cyprus Institute could attract and where they would come from, which was raised by Prof. Koziris. It is intended that the regional area would supplement student numbers at the Institute and the fact that CyI is an English speaking institution should make this a very feasible solution. The second issue concerned the network connectivity both to and within Cyprus which was raised by Dr. Tsapalis. Prof. Kalisperis that, at present, Cyta held a monopoly over the connectivity and options were expensive with limited bandwidth. Fotis Karyannis added that deregulation was a slow process and options were best pursued through organizations such as EUMEDCONNECT2. Prof. Koziris added that demand pushes evolution and aggregating the demand will force the issue. One of the first issues to resolve is high-speed connectivity between sites within Cyprus and then resolve the external issue. This was the model pursued in Greece by GRNET which established the research network backbone first.

4.5. 1st Turkish Users Meeting, 20th February 2009

Venue: Sabanci University, Istanbul, Turkey
Co-organiser: Prof. Ugur Sezerman, Sabanci University

One of the focal points of the panel discussion was the difficulty experienced by Turkish nationals in obtaining a visa to attend events organised in Cyprus and, therefore, the low possibility of fostering collaborations in that type of environment. Prof. Alexandrou (CaSToRC) assured all present that every effort would be made to ensure that visa issues would be organised in advance insofar as it was possible to do so. In particular, the issue would be raised with the organizing committee in respect of the International LinkSCEEM conference in early October.

The question of what the CaSToRC could offer the Turkish research community was also raised. Given that Turkey has a population of some 70 million, a very large research population and invests significantly in computational research (as evidenced by presentations by ITU and ULAKBIM on the day), this was quite a reasonable question. It was emphasized that the relationship is intended to be a symbiotic one, with a focus on collaboration and the expansion of research programs into new and innovative areas. Also there is intended to be a large educational component to activities at CaSToRC which Turkey could both contribute to and benefit from.

4.6. 1st Lebanese User Meeting, June 19th 2009

Venue: American University of Beirut, Beirut, Lebanon
Co-organisers: American University Beirut

The panel discussion focused on what CaSToRC could offer to the Lebanese academic and research community. One of the focal points of the discussion was the educational opportunities CaSToRC will be providing to the Lebanese and general Eastern Mediterranean academic community. The current situation on pursuing PhD studies in Lebanon and specifically at AUB was described and the potential for establishing joint research projects with CyI was discussed. CaSToRC/CyI initial plans for offering a Computational Science and Technology PhD and fellowship program were presented and participants had the opportunity to make comments and provide feedback. All parties agreed on the need of offering regional educational programs of high quality enabling students from the region to invest their lifetime career in their homelands and, thus, reducing the risk of brain drain for the Eastern Mediterranean. The EU willingness to promote joint efforts in Higher Education and the program call for Cross-Border-Cooperation was mentioned as possible sources of support. Ideas about a follow-up LinkSCEEM project, LinkSCEEM II were presented; it was stressed that it should include more participation from countries in the region like Lebanon.

The panellists also discussed the availability of computational facilities in the region and the need for additional HPC systems and a regional grid. The role of CaSToRC as a portal to state-of-the-art HPC both in Europe and the US was presented. CaSToRC representatives also emphasized the willingness of their institution to explore common research interests and urged discussion participants to investigate current research thrusts at CyI. Prof. N. Nassif stated that the Lebanese side will define its needs and interests, prepare a list of research areas and researchers interested in HPC, and communicate with CaSToRC to initiate a discussion.

4.7. 1st Syrian User Meeting, July 6th 2009

Venue: Higher Institute for Applied Science and Technology, Damascus, Syria
Co-organiser: HIAST

The panel discussion initially focused on the actions to be taken following the specific meeting and the completion of the LinkSCEEM project to facilitate regional cooperation and the advancement of scientific computing in the Eastern Mediterranean. Participants urged

for a follow-up project that would take into account the findings of LinkSCEEM. Suggestions were made for the organization of thematic meetings to bring scientists with common interests from the region to contact and, thus, promote scientific cooperation.

Considerable discussion time was also devoted to the CaSToRC Cy-Tera HPC facility with the participants making several comments of interest. It was noted that Cy-Tera has the potential to fill a noticeable void in the region and that general user meetings, such as the one in Damascus, serve the important task of informing interested scientists. The panellists emphasized the importance of developing a reliable HPC facility with good user support for enhancing its acceptance rate by the scientific community. Moreover, they stressed the need for a coordinated educational effort in the form of workshops, training programs and a graduate program in Computational Science. Discussion on the latter topic included references to attracting young talents into computational science and suggestions to develop a scientific computing culture. The participants concluded this part of the discussion agreeing on the need for closer cooperation in educational and scientific programs among institutions from the region.

An additional topic raised related to the problem of network connectivity among the countries of the region and between the region and the rest of the world. The current status was described and the general needs for the future effective use of the Cy-Tera research infrastructure by regional partners were outlined. It was clarified that present connectivity between Cyprus and Syria cannot support reliable usage of Cy-Tera or other data intensive joint research effort and that there is a pressing need for a quick resolution to this problem. Both CyI and Syrian community representatives stressed the need for government involvement to any efforts to resolve the problem. CyI representatives informed about the upcoming project calls funded by the European Union that might be used for initiating a joint project aiming to address the connectivity problem.

4.8. Life Sciences Thematic User Meeting, November 5th 2009

Venue: Larnaca, Cyprus
Hosted within the 9th International Conference on Information Technology and Applications in Biomedicine

Panel Discussion:
Promoting HPC use in computational science in the Eastern Mediterranean region: The case for Life Sciences Research
Panel members:
Prof. K. Schilling (The Cyprus Institute), Prof. D. Fotiades (University of Ioannina), Prof. G. Turkiyyah (American University of Beirut), Prof. M. Abouelhoda (Nile University), Dr. V. Promponas (University of Cyprus)

The panel discussion focused mainly on the Cy-Tera system with the participants showing keen interest to learn more about the deployment and use schedule of the infrastructure. Dr. Abouelhoda and Dr. Zeinalipour (University of Cyprus) requested additional information on the mode of access to the machine. A related question by Dr. A. Ioannides related to the accessibility of the Cy-Tera system to industrial organizations and specifically SMEs. Prof. K. Schilling (CaSToRC/CyI) described in detail the planned resource allocation procedure to be

put in place and the various possibilities to obtain computer time/resources. A discussion followed where Dr. A. Thomas described how industry gains access to computational resources via so-called industry programs at various facilities worldwide.

Dr. Luis Kun, Senior Research Professor of Homeland Security National Defence University/IRMC, USA raised the confidentiality issue of the machine posing questions related to user data protection, network security, etc. He stressed that in the US, and in general whenever confidential data is involved, privacy is a big issue. Participants from the industry also voiced their concerns about the issue and a discussion followed where the methodology used by established HPC centers worldwide with respect to data protection was presented. Dr. D. I. Fotiades explained how the HPC centers he has cooperated with in the past have tackled the topic of confidentiality. He also presented his experiences about efficient mechanisms for resource allocation by HPC systems and effective structure and staffing of HPC centers. Considerable discussion time was also devoted to the needs of the life sciences community with respect to computational resources and the variety of tools that life science researchers use. The discussion was helpful in pointing out the nature and diversity of computational tools used as well as some of their distinct requirements.

An additional topic raised related to the problem of network connectivity among the countries of the region and between the region and the rest of the world. The current status was described and the general needs for the future effective use of the Cy-Tera research infrastructure by regional partners were outlined. Dr. A. Thomas stressed the importance of the issue and its central role in the success of the Cy-Tera facility as an HPC infrastructure of regional dimension. Prof. Turkiyyah pointed out that cables have been installed and are ready for use, but, lack of competition between regional telecom companies and lack of government support through financial or other means prevents the upgrading of regional connectivity to the necessary levels. CyI representatives informed about the upcoming project calls funded by the European Union and how they may be used for initiating a joint project aiming to address the connectivity problem.

The discussion panel concluded by noting that Cy-Tera has the potential to fill a noticeable void in the region and that maintaining communication with interested scientific communities, such as the life science community, is crucial for the success of the project and the general benefit of the region. The participants emphasized the importance of developing a reliable regional HPC facility with good user support in order to facilitate its acceptance by the scientific community.

4.9. 2nd Israeli Users Meeting, 1st December 2009

Venue: Technion-Israel Institute of Technology, Haifa Israel
Co-organisers: Dept. Of Physics, Technion

Members of the Panel discussion:

- Aaron Yaroslavsky, Director of computer center at Technion
- Norbert Attig, Head of the division "Application Support" at Jülich Supercomputing Centre
- Constantia Alexandrou, Chair of Interim Governing Board of CaSToRC
- Pinhas Bar-Yoseph, Dean of Faculty of Mechanical Engineering at Technion
- Simon Brandon, Assoc. Professor in Chemical Engineering at Technion
- Amnon Stanger, Associate Professor of Chemistry at Technion

Three main questions were posed as a seed point for the panel discussion:

- How do we realise a regional education program in computational simulation?
- How can we create sustainable access to high performance computing hardware and software?
- How do we go about promoting regional scientific collaboration?

The panel members then made some opening statements based on attempting to address these questions.
Aaron Yaroslavsky (AY) highlighted that the computing center at Technion had purchased a new Nehalem-based cluster through the implementation of a shareholder system. The shareholder system works on the basis that each user must bring at least 2 nodes to the system and will be entitled to usage based on the resources that they bring. This initiative began in Tel Aviv and has proven to be quite successful.

Constantia Alexandrou (CA) maintained that establishing scientific links is of the utmost importance but that of course the infrastructure is still needed. She proposed that the infrastructure be built on a 'regional' model such as that advocated by PRACE in Europe. Such a system helps address the fact that while computers may be 'cheap' there are large amounts of resources consumed by (inefficient) energy usage. She also outlined that in the proposed follow up to the LinkSCEEM project there would be funding for significant training activities in key areas such as climate modelling, digital cultural heritage and synchrotron research (which are considered to be of regional significance).

Norbert Attig (NA) explained that an organisation such as PRACE exists because cutting edge resources are expensive and have a very short lifespan – the current leadership position head by Juelich is dependent on future funding and only guaranteed for the next 3 years. He highlighted that researchers are frustrated when current computing resources do not meet their needs but must appreciate that there is a learning curve to follow when graduating to larger machines. Computing systems get increasingly faster but this is because of increased numbers of processors not increased processor frequency. This forces researchers to (re)consider the scalability of their algorithms at some stage when following this learning curve. That is why Juelich would advocate a portal approach to aid users in following this learning curve, where they would graduate from local resources to those at CaSToRC to those at Juelich (for example).

Pinhas Bar-Yoseph outlined the reasons behind the closure of the High Performance Computing Center (HPCU), National Inter University Computation Center (IUCC) of which he was Scientific Director. He said there was a lack of government support and understanding which starved the center of resources. The effect of this in terms of his own resource usage is that he must now apply to Juelich for computer time, returning to the situation he was in 10 years ago. He also highlighted that Israel and Cyprus could work together but a two party collaboration is not regional and other neighbours would be difficult to attract into such a consortium.

Amnon Stanger (AS) raised the point that he is a 'user' and not a programmer and is in the situation where he relies on the HPC community to provide him with the tools necessary to utilise these types of resources. CA added that the successor project to LinkSCEEM will

address this by following the learning curve outlined by NA and that the goal is to give people the capability to apply for the highest level of resources.

Joan Adler asked that, as a user, what the next step is for using a machine such as those available at Juelich? She asked whether the situation was different for someone using MPI or a package such as 'Gaussian'? She also asked what the stepping-stone machine in Israel would be? She pointed out that the HPC-EUROPA program filled a void left by the closure of HPCU. NA replied that the entry level for the Juelich BlueGene JUGENE is scalable use to 4000 processors and in the case of Juropa it is efficient scaling to 64/128 processors. Anne Weill asked what the definition of efficient scalability was and NA replied a scaling performance level of 60-80%. Ben Svetitsky asked whether 'trivial' scaling (such as the Monte Carlo sampling that occurs in lattice QCD) counted and NA replied that the facility was not intended for farm computing and must involve communication.

Simon Krichak reiterated that education was important but that he personally wished for a center that could provide archival storage for his experimentation. He has spoken previously to NA that scaling and optimisation were not his area of activity and that he would prefer the community approach where software is selected, ported in an optimal way and then provided to the larger research community.

CA said that scalability training would be part of the LinkSCEEM successor and also commented that access to facilities would be by application and peer-reviewed and that the machine would be available by October 2010. Joan Adler added that HPC-EUROPA would continue to 2012 but that it requires an exchange of scientists and a host must be found.

4.10. LinkSCEEM International HPC Conference, 6-8 October 2009

Venue: Annabel Hotel, Paphos, Cyprus

Chair: Thom Dunning (NCSA, University of Illinois - USA)
Panel Members: Constantia Alexandrou (CaSToRC, Cyprus Institute - Cyprus)
Hafeez Hoorani (SESAME - Jordan)
John Towns (NCSA, University of Illinois - USA)
Pinhas Bar-Yoseph (**Technion - Israel**)
Thomas Schulthess (CSCS - Switzerland)

Panel Questions:

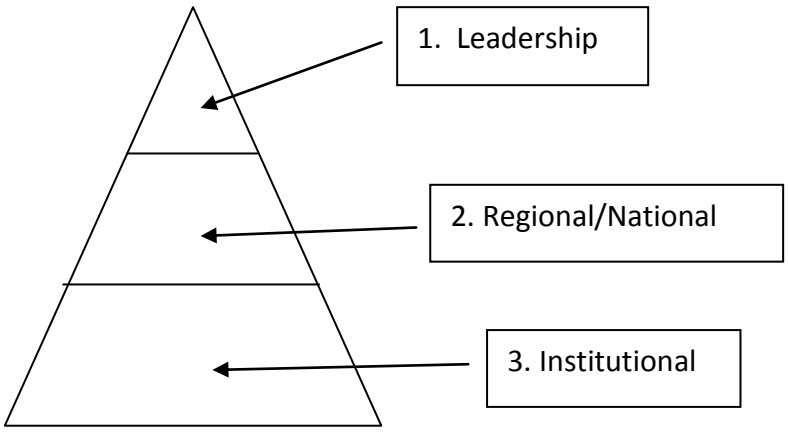
It was intended that the discussions of the panel centered around three main questions:

1. What is needed to advance computational science, engineering and technology in the Eastern Mediterranean region?
2. How can the Cyprus Institute and its Computation-based Science & Technology Research Center contribute to the advancement of computational science, engineering and technology in the region?
3. How can the region position itself to benefit from the major supercomputing initiatives underway in Europe, the U.S. and Japan?

T.Dunning opened the discussion by posing the above questions to the panel members. P.Bar-Yoseph briefly discussed his own experience in establishing an HPC Center in Israel where he acted as Scientific Director for the High Performance Computing Center (HPCU) at the National Inter University Computation Center (IUCC). He noted that while initial funding for the center existed, longer term funding was expected to be provided by the University but the University finances ultimately could not support this. The center had a good initial placing in the Top500 but by 2004 was out the list by 2 orders of magnitude. In that period the centers hardware was rapidly surpassed by technological developments.

A burning question for an emerging center is then how to improve the 'gain-to-pain' ratio? In addressing this question one really needs to address (some of) the "Top 10 Challenges in Parallel Computing" as written by Micheal Wrinn of Intel. Doing so would allow a bridge to form between the center and the wider HPC ecosphere (mediated by organsiations such as PRACE) and reduce the impact of hardware aging.

H.Hoorani added that training for large scale computing facilities should not be focused solely on young students but should also address mid-career scientists in the region since they have the capacity to create research initiatives in these areas. This would also facilitate human networking which would be an essential part of the proposed community. It would also address the fact that people in the region, in general, really lack the essential knowledge on how to use HPC.



T.Schulthess reiterated that computational simulation is indeed the third pillar of science. He stated that he believed that simulation requires a tiered HPC ecosystem (opposite).

An important *application* example of how this structure is used is that of VASP. It went from being an institutional to a national to a leadership application and **back** to a national application. What is important is not the computational power of the facilities but the simulation capabilities of those facilities. This requires a spectrum of capability computers and we need to ensure that applications move up and down the ecosystem and that the spectrum of capabilities is utilized. He would recommend that a factor of 5 should differentiate the tiers in this system (1.5PFlop/s -> 300 TFlop/s -> 60 TFlop/s). In 2011/2012 it is expected that the top tier would move to 20 PFlop/s which would mean an institutional level of 800 TFlop/s.

In direct response to the questions, T.Schulthess felt that a connected regional ecosystem must be built and it should be built in such a way that the connection path to larger systems is well defined. In response to the second question, he felt that it is reasonable to create a regional center. He believed the focus should be on enabling technologies and that computer science and mathematics should be nurtured in this environment leveraging some of the experiences of the KAUST center. He advised to act independently and to cherry-pick

the focal areas. In terms of architecture he would recommend a commodity-based node architecture with GPU accelerators. To the third question he advised an indirect addressing of the issue, find first what your strengths are, build on them with the viewpoint that computational simulation is the goal. As a result opportunities worldwide will present themselves if the science is good enough.

C.Alexandrou agreed that a pyramid structure is indeed appropriate and that the access path to the top tier must exist. The partners that CaSToRC have are experienced and can make that path both possible and practical. She found that LinkSCEEM was a good learning process with the next step being to promote active collaboration in areas which are of particular interest across the region, such as climate change and digital cultural heritage. In particular, she noted that we need to assemble the resources to make this effort sustainable:

- need to be coupled to developments in Europe,
- need to promote appropriate educational programs,
- need to build on human capital.

J.Towns began by saying that while there is clearly a great deal of interest in the region there is a lack of organization and LinkSCEEM needs to be persistent. A solution will present itself by building a community mad CaSToRC can act as a magnet for that community. Yes, there is a need to engage the global community but there is a need for cohesion in the region first. The real question is: what are the strengths of the regional community that we can offer to the global community? In that sense the region should do something exciting that can give a focal point to a regional effort.

At this point, T.Dunning opened the floor and invited questions and comments from the audience.

R.Ahmad of SESAME wondered whether there a computing advisory committee for the region should be developed to aid in building the tiered structure.

V.Jongeneel of the Cyprus Institute agreed that an ecosystem that flows both upwards and downwards is appropriate and that we should define our own strengths and needs rather than solely adopting those from the top tier.

Another person from the floor raised the point that many people are still not aware of the ongoing LinkSCEEM activities and that workshops would be very welcome. T.Dunning said that such things are part of the outreach activities of the project and efforts in this direction are ongoing.

N.Dalfes of the Istanbul Technical University commented that everyone will build their national effort but it was establishing the pyramid that was difficult. There is a need to empower people at the different scales. It is also a question of culture as well as ecosystem; we need to cultivate people to experiment in their own area. Turkey's efforts involve workshops twice a year but this is not enough. The region needs a forum to share experiences and knowledge (such as power requirements and reliability of hardware). Thematic efforts are useful and high performance computing is the key element. V.Jongeneel commented that such ideas are exactly along the line of thinking for

LinkSCEEM2. T.Schulthess advised to spend time on deciding what is needed in the region and develop an understanding of this so as not to repeat the mistake that others have made.

N.Nassif of the American University of Beirut commented that education is key to the questions posed. There is a need to establish sustainable curricula in these areas that are acceptable to academia. Such a thing is difficult in simulation science because it is so multidisciplinary. Because of this initiatives to properly establish educational programmes are needed.

R.Ludijten of IBM Research Lab Zurich advised to ensure that the human capital is developed to run the center in 10 to 20 years. People should be given the space to let the center define itself and find its own identity. C.Alexandrou commented that to enjoy the long term success of centers such as NCSA or Juelich input is needed from many people. To this T.Dunning added that we should also try to learn from the *mistakes* that these centers have made.

To conclude T.Dunning made a few final remarks again addressing the three questions posed. He said it is not a lack of academic problems that heeds advancement. Many exist and of these many have regional scope. If activities can be correlated across borders, that would lead to an unrivalled resource. He advised not to get computer envy but to focus on solving problems.

As a final remark he reiterated the need to build human capital. He advised that colleagues and students are the greatest asset you have. He also commented that education is indeed critical. In particular he advised that it is now time to stop teaching serial programming, parallel programming is the future regardless of how large the machine and the time to focus on teaching that is now.

Discussion

In summary, it is clear that the user meetings have been a major component of the regional networking process. Through the presentation to potential users of the LinkSCEEM project and its context, the presentations by users of their research activities, resources and needs, and the dialog occurring in the round table or panel discussions, they have greatly favoured a two-way communication that is proving extremely useful both for gathering knowledge about the future CaSToRC user community, and for channelling information to its members. Both aspects will contribute to the community's build-up, notably by creating strong links with CaSToRC, and a sense of involvement and ownership from users who will realise that their input is taken into account for the design of the future CaSToRC HPC facility, and for the related research and training programmes.

The panel discussions have clearly outlined that computational resources alone are not required by the region. Indeed in a large portion of the discussions, hardware resources were not even mentioned. Focal areas repeatedly included education and training and, in particular, a coupled improvement of network connectivity. Only these could address the

accessibility concerns of the wider community and the long term sustainability of the overall effort.

The User community itself must appreciate what its own resources are. These are not just computers, but people and the characteristics of the region. Identifying research areas where these regional strengths are best utilised, and where a global impact can be made - both because of their regional relevance and their potential international visibility - and focussing on these areas can contribute to the sustainability of the entire process.

5. Assessment of Regional HPC Resources

A preliminary survey of existing HPC resources in the Eastern Mediterranean had been carried out at the beginning of the project, resulting in a document describing the existing HPC infrastructure in the region; an intermediate version (see Annex 1) is available.

During the preparation of the LinkSCEEM-2 project (which can be considered to be the implementation phase of the LinkSCEEM project), updated information with respect to this report has been obtained.

The estimated world capacity as reported by the June 2009 TOP500 list is about 34 Pflop/s. Only two countries in the Eastern Mediterranean area have computers that appear in this list, namely Israel and Saudi Arabia with a combined 0.6% of the total capacity. Since simulation has come to be regarded as the third pillar of science (the others are theory and experiment) one may take it as another measure of how advanced scientific research is, and unfortunately according to this measure the region is rated well below Europe and the US. The lack of resources is one element that hinders the development of user communities.

Others, and perhaps even more serious ones, are the lack of expertise and the limited integration of the scientific communities, so that the build-up of HPC infrastructure must be paralleled by the development of adequately trained human resources, both for technical and scientific personnel. The integration of the scientific communities must parallel that of the resources, and be supported by networking activities, and accompanied by a general raising of the awareness of the importance of HPC and computational science.

It is intended that three regional HPC facilities will contribute resources and services to the LinkSCEEM-2 project. The project aims at providing activities that go beyond the services that these facilities offer to their institutional and national users, and are aimed at engaging virtual research communities at the regional scale.

Aside from the facilities of CaSToRC, the Bibliotheca Alexandrina (BA) has recently installed a SUN cluster of peak performance exceeding ten Tflop/s, and also has an extensive archival system. In addition, its visualization infrastructure, incorporating an active stereo immersive environment, is one of the most advanced in the region, and the BA conducts joint activities and partnerships in visualization projects with Universities in Egypt, the region and internationally.

Besides CaSToRC and BA the HPC facilities at NARSS, currently consisting of a one-rack Blue Gene/L, are expected to be upgraded to meet the needs of the larger user communities that will be developed. Its resources also include a visualisation lab, a satellite station and facilities for remote sensing and satellite image processing.

Additional sites are expected to be developed and integrated into the LinkSCEEM2 project, such as the computer resources that are being planned for on site processing of SESAME data, as well as the computer resources of several Israeli institutions.

6. SESAME – Assessment of Needs

Rami Ahmad of SESAME has conducted an extensive study into the computing needs of SESAME. This report is attached as Annex 2 and represents the outcome of his visit to research centers in Europe for a better understanding of the computational requirements of the SESAME scientific program. The discussions contained therein are based on the computing questionnaire he prepared for those visits.

It was found that the most computationally demanding beamlines are Protein Crystallography, SAXS, and Tomography beamlines. These cannot afford to be on remote compute resources or queuing systems since interactivity is a very important factor. Since the online computing resources demands are so high, local high performance computing resources are required at SESAME. A computational power of 11 TFLOPS is estimated to be sufficient to accommodate the online requirements of the complete SESAME scientific program (phase-I & phase-II) with any further expansion in number of beamlines which is foreseen in future. The remote computing power allocated by CaSToRC is needed to accommodate the beamline users' analysis and simulations requirements where the computing power needed by these users is equal or can be greater than the online computing power available at SESAME. In addition the remote HPC resources can be used as a scalable or a fail-over HPC to the local HPC facility.

It is expected to have a large amount of generated experimental data from the beamline detectors, estimated to be of the order of a few TB per day, based on current cutting-edge detectors technologies. This number is expected to increase in the future based on evolving technologies in beamline detectors. Local data storage capacity of 100 TB scalable to 500 TB and an archival system with a capacity of 1.5 PB to keep the experimental data for a period of one to five years is required. The experimental data is non-regeneratable and precious so a local datacenter with disaster recovery is also required.

For the SESAME computing model to work, where HPC facilities are remotely based in the Cyprus Institute, a high throughput network connection between SESAME and CyI ranging from 155 Mbps to 1 Gbps is necessary to transfer the experimental data efficiently. Beamline local network preparation with 1Gbit/s Ethernet is needed as well, fiber connection with bandwidth 10Gbit/s or Infiniband access is required between the HPC cluster and the local data storage. The needed skills to build and operate the beamlines computing infrastructure are: two system administrators (Linux and Microsoft Windows), one network engineer, one scientific software programmer and four hardware technicians.

7. Conclusions and Outlook

A. Assessment of needs:

The main conclusion of this report is of course the existence within the Eastern Mediterranean of significant interest and needs for computational resources, and for related research and training programmes; this has been clearly demonstrated during the networking process, notably through the simplified survey and during the presentations and discussions that occurred in the user meetings. This constitutes an important validation of the concept of CaSToRC as a regional HPC facility and Center of Excellence for computational science.

The results described above will provide useful input for other Work Packages of the project, notably for the preparation and incubation of research and educational collaborations (WP4) and the Network Study (WP5). Indeed many opportunities for collaborative research can be derived from the content of the user database, from the simplified survey and from the user presentations in the various meetings. In addition obvious needs for training and education are apparent from the users surveys, since whilst there are certainly a number of users proficient with parallel computing, the majority are lacking basic skill-sets in this area.

Education must therefore be a core component of the CaSToRC's development; it should take a number of forms, and even a tiered structure, to adapt to the different skill levels already existing in the community (see the WP4 report for more detail):

- Education in community-maintained and commercial software packages and tools
  For a large number of users, package software is already available to address their research needs and does not require major programming contributions on their part. An educational program that informs Users of such software, and how it is used, could greatly benefit users in the short term and is a highly appropriate first step for those unfamiliar with HPC environments.
- Education in parallel programming models and tools
  For those who already have programming experience, but not necessarily in a parallel environment, it would be useful to have introductory courses on parallel programming models, concepts and tools. These might include, for example, an introductory course in MPI and in the use of optimisation tools.
- Education in emerging HPC technologies and tools
  For those familiar with a parallel environment there is still scope for further education. Following the example of PRACE, one can learn of emerging parallel specific languages such as Chapel, Fortress or X10 as well as programming techniques for accelerator technologies like GPUs and FPGAs.

A related question is how the users' perception of their own needs may evolve in time, in parallel to the increase of their resources and expertise; it will be of interest to monitor that evolution, and perhaps to get insight by evaluating the said needs before and after taking part in a training programme, as has already been done in a preliminary fashion for the small scale precursor training efforts that are mentioned in section 3.

It is also worth mentioning that the Computer Science community, interested in large systems, large databases, etc. from a theoretical or practical computer science research perspective, could make a very valuable contribution to future training and advanced user support activities that could gain significantly from an active interface with basic IT/Computer Science research. This component should be kept in mind in relation to the incubation of the research and education collaboration.

It is also clear that CaSToRC must actively introduce the regional scientific community to HPC. It should not only provide an educational structure that addresses all levels of experience, as mentioned above, but also create a user policy and process that encourages usage, particularly for new users. In this respect the setting up of a specific access channel for very small needs, with a simplified application process and enhanced user support where needed (e.g. for new users) could be very favourable.

It is worth noting that the considerations in the two previous paragraphs illustrate the interesting complementarity of the surveys: potential new HPC users certainly exist within the participants of the simplified survey, while insight into the interface between advanced user support and research on emerging HPC tools can be provided by the detailed survey.

Concerning the Network Study, there is no doubt that the assessment of needs process has confirmed its relevance, since network connectivity has emerged as an obvious concern for the scientific communities in the whole region (only Greece, Israel and possibly Turkey have a different status in that respect, although scientific communities in these countries are also interested in regional connectivity). This topic was stressed in almost all Users Meetings. It seems that the consensus among users is that this problem is more political than technical. The solution requires both large investments by governments and cooperation among them. Even if financing could be found to lease more bandwidth, connecting through Israel will be problematic for several countries.

In addition the outcome of the assessment of needs, and notably the analysis of the detailed survey, has also provided interesting input for the design of the future CaSToRC facility. At the current preliminary stage, it seems that the user surveys indicate that a tightly coupled Linux-based cluster is the system of choice for the majority. This would also be a very prudent choice given that such a system is the most prolific worldwide and a great deal of support and software is publicly available, making the transition to such a system easier for novice users. Accelerator technologies can also be built into such a system meaning that there will still be plenty of raw power to satisfy more advanced users.

According to the development plan of CaSToRC, a tier-2 machine will be implemented in 2009-10, followed by a much larger tier-1 machine within the next few years; having this type of tier-2 facility will provide a highly appropriate breeding ground for advanced users that can then utilise the necessarily more complicated structure of the future tier-1 facility. It can also, given its cluster nature, help address some of the capacity computing requirements of the regional user community.

B. The User Community:

The current version of the database provides a picture of the user community; it seems clear that the setting up of the database and the carrying out of the surveys have been assisted dramatically by the holding of User Meetings. Participation in both of these activities increased significantly once direct contact had been made through the User Meetings. The Amman user meeting is an interesting case, however, where there were a large number of additions to the database without any contributions to the user surveys. This highlights a certain reluctance for exposure in some countries but also a definite willingness for participation. This is evidenced by the repeated participation of Jordanian institutions in outreach activities and the holding of the 3 day HPC workshop in Amman. As mentioned above, the specificities of each community should also be taken into account in the planning of future activities.

The database will also serve for the follow-up to potential users and as a basis for future actions. As mentioned above, it appears that, at least in certain countries such as Israel, an enhanced networking effort, carried out at grass root level and favouring direct contact with end users would be in order. This should furthermore be adapted to have significant impact on Graduate students and young researchers, who turn out to be very receptive if approached adequately, as was shown by the tutorial day that was held within the LinkSCEEM International Conference (see sections 3.3 and 4.10). For certain scientific fields, thematic meetings can be an appropriate scheme, and the organisation of LinkSCEEM sessions in community events has turned out to be a very interesting solution, for instance for Life Sciences (see section 4.8).

Follow-up activities will be carried out in order to inform the contacts that are identified within the database about future LinkSCEEM events and activities, about the development of CaSToRC and about other relevant issues. Contact and coordination will be ensured with the members of the user contact groups, in order to involve them in relevant activities.

Direct contacts between users should also be facilitated, possibly through the setting up of an adequate interactive online platform.

In the future, the networking process initiated within LinkSCEEM will enable the gradual development of an incubation process that will produce concrete collaborative research and educational thrusts; these will act as magnets and facilitate further networking and engagement of the user communities. The continuous monitoring of user needs will remain necessary, and can be carried out in parallel to this ongoing networking process.
The LinkSCEEM-2 proposal, if funded, would be a major step in that direction, including both networking, integration of e-resources as well as collaborative research and training, and focusing on climate science, cultural heritage and synchrotron applications, three fields of major importance, both globally and in the region.