The SNIC User Survey 2014

Editor:
Johan Raber

National Supercomputer Centre
Linköping University

December 11, 2014

Contents

1 Introduction

This report summarises the responses to the "SNIC User Survey 2014", surveying the opinions of users and maintainers of the SNIC facilities. The number of respondents of the survey was 159, representing most HPC utilising scientific communities, experience and academic levels as well as SNIC Centre staff. There were 8 responses from SNIC center staff members.

A certain degree of editing of the responses to the free format questions was unfortunately unavoidable. Most edits have been done in terms of formatting but not in terms of content, while some content have been "blanked" so as to not reveal names or otherwise identify people. This pertains to survey respondents and others, not any SNIC staff who are mentioned in the comments.

Responses to free text are uniquely tagged by respondent, the numbering based on the order in which respondents have submitted their response. This corresponds to their row number of the form response spreadsheet published at URL `https://www.nsc.liu.se/support/Events/SNIC_user_forum_2014/user-survey`, which contains the raw survey data. Editorial comments are set in bold and enclosed in square brackets throughout this report wherever it may be unclear from the context that it is an editorial comment.

Please note that the question numbering in this report does not always strictly follow the order they appeared in the survey but in those cases follow the order in which they appeared in the raw form data spreadsheet as presented in Google docs. The question numbering directly maps to the column number in this response spreadsheet. Comparing the questions to the fields of the form data, you will also find that some fields of the raw data are left out in this report. These fields either contained no data or are of very little interest in this context, for instance field one contained the time of form submission for the respondent and was left out.

2 Respondent Data

Q2: *Your Name*

[**Responses are left out the report and the responses spreadsheet**]

Q4: *How many years have you been using SNIC resources (HPC and/or storage)?*

Years usage experience	Ratio
More than 3 years	60%
Between 1 and 3 years	26%
Less than 1 year	12%
I do not personally use SNIC resources	2%

Q5: *How frequently do you use SNIC resources?*

SNIC resource usage frequency	Ratio
Daily	51%
Weekly	21%
Monthly	17%
Less than once a month	6%
I do not personally use SNIC resources	4%

Q6 – Q14 polls the individual center usage within SNIC and had three levels of usage to choose from "NO usage", "SOME usage" and "MUCH usage". The figure leaves out all answers in the category "NO usage" and the responses in the category "I do not know/No answer/Not applicable" as it dwarves the other categories and it is arguably more interesting where people do use systems than where they don't. The answers are shown in figure 1.

The original form input request was: *Mark the SNIC centers where you have used SNIC compute or storage resources in the past 24 months*

Q6: *PDC,* **Q7:** *NSC,* **Q8:** *HPC2N,* **Q9:** *C3SE,* **Q10:** *UPPMAX,* **Q11:** *Lunarc,* **Q12:** *I do not know,* **Q14:** *No answer/Not applicable*

Q15: *Have you applied for, or do you have access to, other computing or storage resources for your research during 2012-2014?*

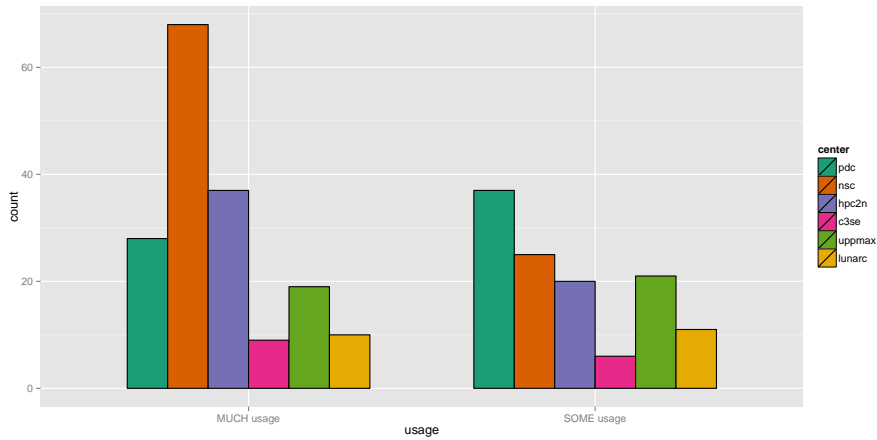| Have other compute resources	Ratio
Yes	30%
No	70%

Figure 1: User reported SNIC center use.

Q16: *Which compute or storage resources outside of Sweden do you have access to or have applied for?*

    R5:
     "PRACE, ECMWF"

    R10:
     "PRACE, ECMWF"

    R11:
     "ECMWF"

    R15:
     "No"

    R16:
     "Xsede, futuregrid"

    R17:
     "PRACE,INCITE"

    R26:
     "INCITE; access to wide range of machines for testing Gromacs portability and performance"

    R30:
     "ecmwf"

    R33:
     "PRACE, XSEDE"

4

R34:
  "prace"

R35:
  "LRZ "

R44:
  "PRACE"

R45:
  "PRACE, NSF/XSEDE"

R46:
  "Anton, XSEDE"

R53:
  "PRACE"

R56:
  "none"

R61:
  "we have our own calculation server"

R64:
  "Broad institute"

R74:
  "Dawning"

R75:
  "PRACE"

R77:
  "PRACE"

R84:
  "applied for PRACE"

R86:
  "HLRN, Germany"

R88:
  "PRACE"

R93:
  "AWI computing center (Alfred Wegener Institut, Bremerhaven)"

R103:
  "Local Resources France - Limoges"

R104:
  "CURIE"

R107:
  "PRACE DECI-11"

R112:
  "NSF"

R119:
  "EGI"

R120:
  "WLCG"

R125:
  "Sisu (CSC supercomputer)"

R126:
  "Juqueen"

R128:
  "HPC of China"

R142:
  "Molecular Foundary (Staff coll+ on site CPU)"

R144:
  "Not outside of Sweden but the SMHI computer resources at NSC"

R153:
  "PRACE 2012"

R154:
  "PRACE"

R158:
  "PRACE/DECI (PDC)"

Q17: *Employer*

Employer	Ratio
Royal Institute of Technology	25%
Uppsala University	14%
Linköping University	12%
Stockholm University	11%
Lund University	9%
Chalmers University of Technology	7%
Umeå University	7%
University of Gothenburg	3%
Swedish Meteorological and Hydrological Institute	2%
Karolinska Institutet	2%
Nordic Institute for Theoretical Physics	1%
Karlstad University	1%
Linnaeus University	1%
Luleå University of Technology	1%
Swedish Institute of Space Physics	1%
Swedish University of Agricultural Sciences	1%
University of Borås	1%
Other	3%

Q18: *E-mail address*

[Responses are left out the report and the responses spreadsheet]

Q19: *Employment Position*

Employment position	Ratio
PhD Student	32%
Post Doctoral Researcher	11%
Lecturer	4%
Docent	3%
Associate Professor	7%
Professor	15%
Researcher	12%
Undergraduate student	1%
Other	15%

3 SNIC Computing Resources

Question: *As a user of SNIC resources, what sizes of jobs do you mainly run and how frequently?* The answers rate the alternatives (Q20 – Q25) on a scale with levels "Never", "Seldom", "Sometimes", "Often" and "Very often". The results are summarised in figure 2 as the ratio of respondent answers in percent.

Q20: Serial, single-threaded. **Q21:** Single-node, multiple cores. **Q22:** Multi-node, less than 256 cores. **Q23:** Multi-node, 256 to 1024 cores. **Q24:** Multi-node, 1024 to 10 000 cores. **Q25:** Multi-node, more than 10 000 cores.
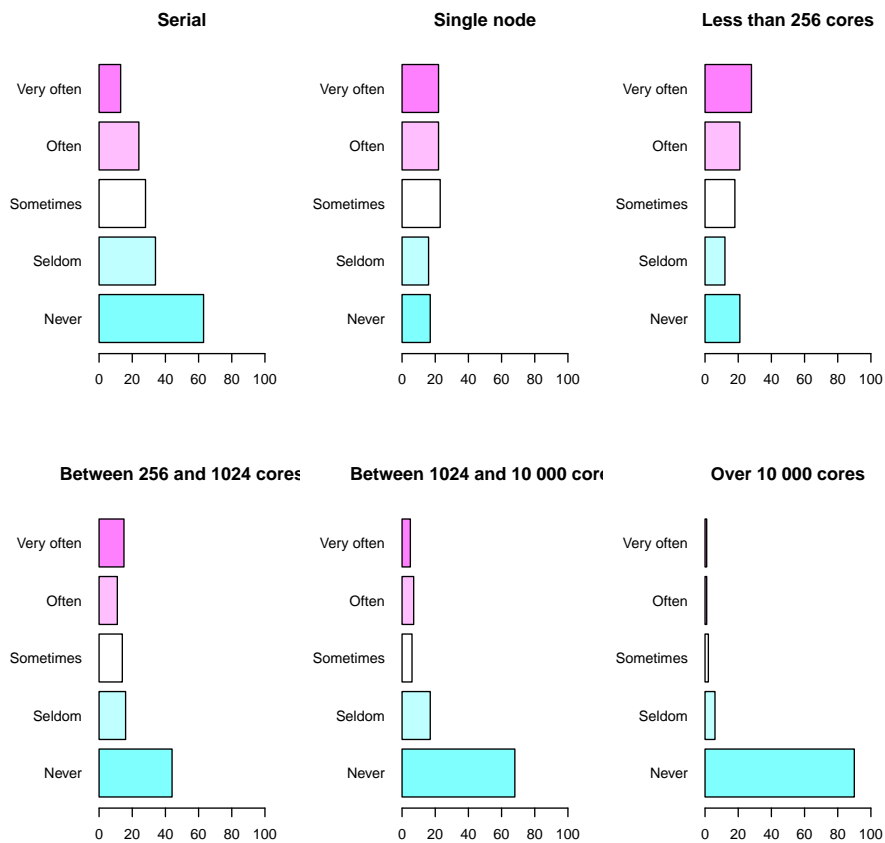

Figure 2: Predominant user job sizes and the frequency with which they are run, counted in number of responses.

Question: *As a user of SNIC resources, how long do you run jobs of various sizes?* The answers rate the alternatives (Q26 – Q31) on a scale with levels "Not at all", "Below 1 day", "Between 1 and 3 days" and "Over 3 days". The results are

summarised in figure 3 as the ratio of respondent answers in percent.

Q26: Serial, single-threaded. **Q27:** Single-node, multiple cores. **Q28:** Multi-node, less than 256 cores. **Q29:** Multi-node, 256 to 1024 cores. **Q30:** Multi-node, 1024 to 10 000 cores. **Q31:** Multi-node, more than 10 000 cores.

Q32: *Is the geographical proximity to SNIC computing hardware important to you?*

Is geographical proximity important	Ratio
Yes	14%
No	86%

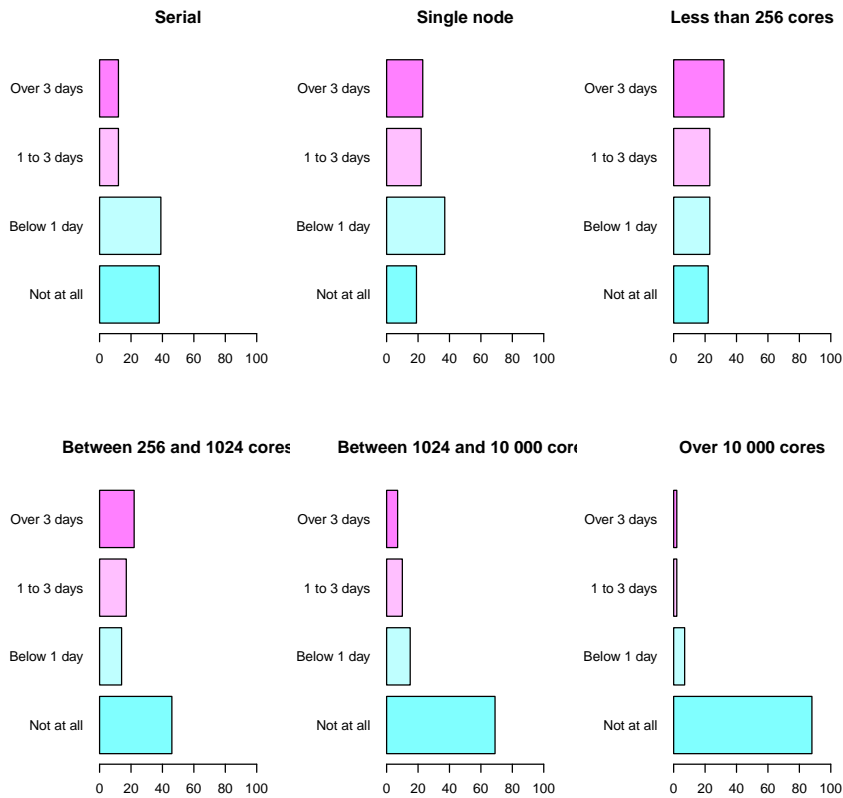**[Please find the motivations for answering "Yes" under Q34]**

Figure 3: Predominant user job sizes and their typical walltime, counted in number of responses.
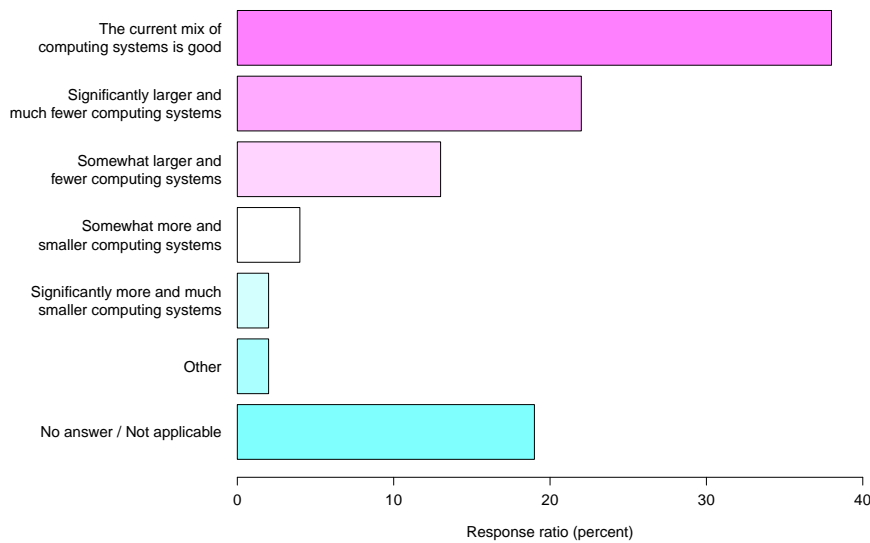
Figure 4: The most beneficial future evolution of SNIC compute resources according to survey respondents.

Q33: *What future evolution of SNIC computing resources would benefit your research the most?*

[**The answers are summarised in figure 4**]

Q34: *The geographical proximity to SNIC hardware resources is important to me because ...*

R10:
"it allows for frequent and rather spontaneous meeting with the staff at the computing centre"

R13:
"A large part of the production calculations can be performed remotely with Little significance of geographical proximity.However, there are significant benefits from having local facilities and local experts to facilitate resolution of special hard- and software issues, as well as to expedite high-priority jobs and questions in an efficient manner. "

R19:
"1. It's easier to get support from a local computer center than from a national center.2. Is threshold for new users is smaller on a local computer center than from a national center.3. The personal at the local computer center give courses on how to use the computer4. The local computer center is used is in quite a number of MSc courses3"

R36:

"I am also in charge of hardware for Abisko, so I am a special case ;) Otherwise, it is often easier/faster to talk to the people in charge of the hardware resources directly"

R48:

"It's not really the proximity to the hardware, but to the staff handling it. The LUNARC team is sort of attached to my department and the close proximity to them has been very fruitful for me. Hence, I like to run on their hardware as I can speak them about it directly in person. Naturally, the LUNARC machines are then located near me. "

R52:

"The support persons are geographically near the hardware, and I can more easily begin and keep alive a relationship with them, if they are in my geographical neighbourhood.As a system expert, I can more easily inspect and handle the hardware if it within walking distance."

R57:

"I work at the NGI and we transfer a huge amount of data to UPPMAX on (what I believe to be) a fairly high bandwidth direct connection. Not sure of the details here though."

R59:

"Possibility to affect hardware and softwareClose to SNIC personnelPossible to extend the cluster by financing additional nodes, that are added to the cluster and made available to me in the for of time."

R70:

"Since we are a production facility which transfer large amounts of data to the UPPMAX cluster on a daily basis geographical proximity is of importance to us."

R76:

"It has allowed me sometimes to be in personal contact with application experts that have been very useful to me during the developing of the program for my PhD."

R105:

"I like to able to talk to peole handling the machine face-to-face.There is also less risk for network failure."

R118:

"Mainly speed"

R119:

"proximity to local experts"

R142:

"Let me first qualify: 'geograophy proximity on one of several facilities is important' to me because I develop functionals and runs all sorts of variants in the development work. I therefore end up compiling a lot as we make code extensions and we benefit from working with a local team that understand why we still need to get the compilation optimized as we go along. We get that service at C3Se (but it should be noted that we are also grateful for hpc2n where we are also getting that kind of help).In addition: we often run computational work in undergraduate projects or with High-school reach-out projects (where we get them to also publish)."

R145:

"Special software procured for use at specific center"

R150:

"I am not the main user, but my PhD student [**name removed**] has had good personal contact with the service people at C3SE, which been valuable for getting the models running efficiently."

R154:

"It is fruitful to be able to meet support in person at some instances."

R158:

"The direct contact with support for porting software and similar, for local workshops, and a good understanding of the local needs and occasional special requirements. "

R159:

"In principle we can use computers on all centers. However, the contact to the local center is much easier and more convenient. It is simply the fact that I can walk over to c3se when I have some problems and solve it in a few minutes instead of several emails back and forth withoout moving the issue forward."

Q38: *The SNIC computing resources typically have 32-128 GB memory per node. Do you need access to compute nodes with larger memory configurations?*

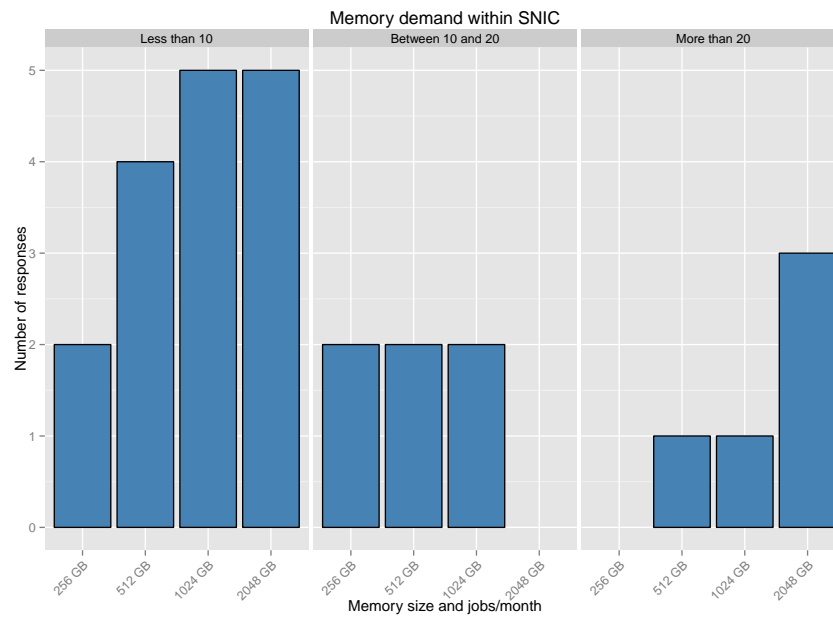| Need large memory systems	Ratio
Yes (please specify how much on the next page) [**See fig. 5**]	19%
No	59%
I do not know	14%
No answer / Not applicable	9%

Figure 5: The memory demand among respondents answering "Yes" to Q38. The panels represent how many runs are estimated to be needed per month and the x- and y-axis show the quoted memory size needed and how many respondents want this, respectively.

Q39: *What size of memory do you need nodes equipped with?*

[See figure 5]

Q40: *How many large memory jobs do you estimate you would run per month?*

[See figure 5]

Q35 – Q37 are of the voting type where respondents were asked to answer the question: *How should the set of SNIC computing resources evolve in the future to benefit you/your group the most?* by rating the three alternatives "Throughput computing", "Capability computing" and "Heterogenous computing" using the rating levels "More emphasis", "Same emphasis", "Less emphasis" and "No opinion". The answers are summarised in figure 6.
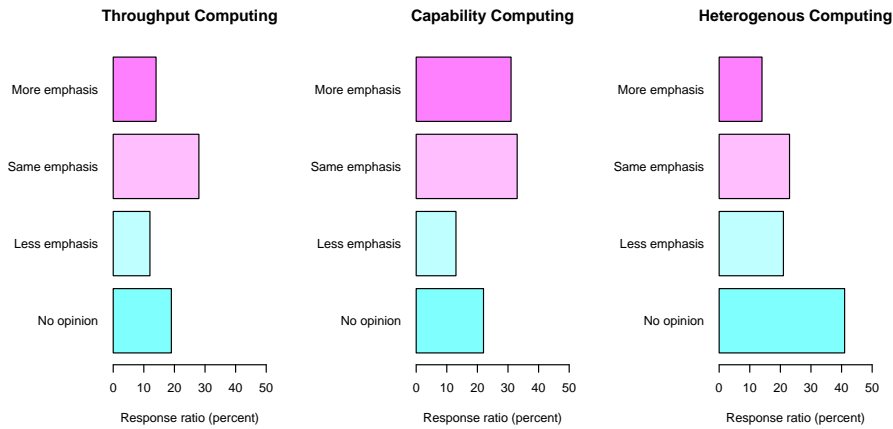
Figure 6: Respondent votes on which future evolution of SNIC computing resources would benefit them the most.

Q41: *In case you have access to multiple SNIC computing resources (possibly at multiple centers), how do you perceive the system configurations/setups on these resources?*

SNIC system perception	Ratio
The systems are configured the same or similar	19%
The systems are configured differently. It does not cause problems for my research	25%
The systems are configured differently. It causes problems for my research	15%
I do not know/ I do not have access to multiple SNIC systems	22%
No answer/Not applicable	19%

Q42 and Q109 – Q110 are ratings based on the input request: *Rate how beneficial it would be for your work if there were common system configurations on the SNIC resources that are of comparable nature.* The answers rate the alternatives on a scale with levels "No opinion", "Low benefit" and "High benefit". The results are summarised in figure 7 as the ratio of respondent answers in percent.

Q42: Login procedure. **Q109:** Job submission procedures. **Q110:** Software selection, availability and naming. **Q111:** Job scheduling behaviour.
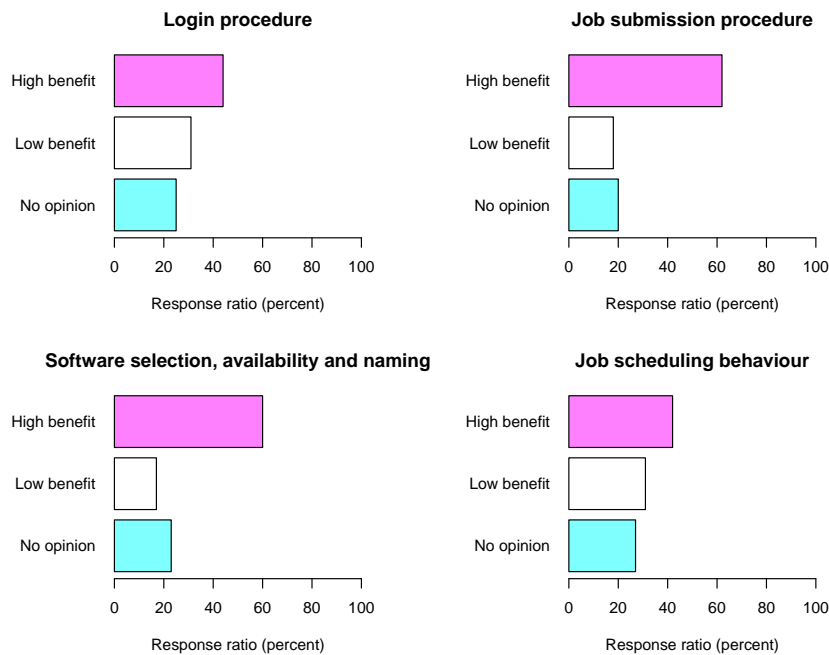
Figure 7: Respondent perceived benefit of increased commonalities between SNIC centres.

Q43: *Please provide any further comments that you would like to add concerning the SNIC computing resources. In particular in case you have expressed your dissatisfaction in any of the answers above, we would be grateful if you could provide more detail.*

R11:
"NSC has set up their system as a very user-friendly system. For example PDC is not as user-friendly (kerberos, etc.). For instance at NSC one invokes the MPI compiler/libraries by the simple flag -Nmpi for ALL different compilers, which makes use very much simpler than at e.g. PDC."

R23:
"Add more licence for matlab parallel tool box"

R24:
"The size of the GPU/heterogeneous clusters is too small and the amount of them is too low."

R25:
"I usually use fat nodes and other nodes using only half of the capacity to increase the memory per processor. It would be good to have more of them, as it would not be necessary to use the job time in hours not effectively

used. This is also true for computing time.My use of PDC is superficial, it's a long time that i don't use for work (heavy usage) purposes, but the NSC login and system usage is much better!"

R26:
"Learning the quirks of multiple

- hardware setups
- login procedures
- file system layouts
- module naming
- procedures for compiling
- accessing interactive nodes
- job script construction
- job submission and
- job management

consumes a large amount of human time.If there are multiple kinds of machine+infrastructure, many people will not migrate their work to best fit the machine capabilities if that means they have to re-implement most of their workflows.In particular, the mismatch between NSC and PDC facilities means I only choose to use one of them at a time. The large range (say, of queueing systems) even within PDC machines is a barrier to using them, and I presume is also a problem for the staff to document and support.Expert-level deployment of conditional login, setup, compilation and job management scripts can defray much of this cost, but this is done on a per-user basis, and is thus very inefficient even if it is done. I think there is a clear role for SNIC to push hard for more uniformity across the whole SNIC compute infrastructure. Any non-uniformity should be required to deliver a clear benefit to end users or support staff."

R43:
"I work very irregularly with the computing centers and keep forgetting the details. I am very happy when I memorise some details and it becomes a hurdle to switch to another center because of this. I also see this with my students. In fact, some of them are worse and can prefer using a center with a minimal quota and possibly longer waiting times, simply because they feel it is too much work to figure things out at another center. It would be great, if it was possible to simply move a well-tried script from one environment to another and know that it will work."

R44:
"You have missed the most important part. The architecture of the systems.

Those AMD based systems are pathetic both in terms of running jobs and memory management. Intel based systems are much better. No wonder that triolith is the most popular system among the lot. I also liked the research idea of NSC before they decided on which architecture they will invest for Triolith. I remember they tested some of the intel based and AMD based systems for the popular software that run on NSC."

R47:
"We would like to use a somewhat different approach using open source storage systems as Apache Cassandra or Hadoop for storing and processing time series data. These databases would, in SNIC terms, be rather small. The important part is that SNIC could possibly provide competence/support around these products, a competence that are hard to find in Sweden today."

R52:
"SNIC centres invent and improve their computing environment now and then. These changes are usually of substantial value for users and would be impeded by a common system configuration."

R53:
"how long do you run jobs of various sizes? This question is unfair. Because of the Q system I run ~ 13:59 jobs so they get through and I can have control/see errors, progress, etc. After 14h I simply restart the same jobs, so the total job time can be big. Very big jobs are set for a day after some testing.In case you have access to multiple SNIC computing resources (possibly at multiple centers), how do you perceive the system configurations/setups on these resources? I say stupid misconfigurations, like banning tcsh shell, different routines, etc. causes discomfort and unnecessary confusion. But scientifically there is no problem with different computers - if I can eventually compile - than it works much alike, despite the inner differences."

R54:
"The current emphasis on number crunching is too narrow. Virtualisation of resources both for temporary analyses and permanent public applications would be beneficial."

R59:
"We install ourselvesIt is better if each cluster makes that cluster as good as possible instead of making everything similar everywhere.We do however like the thinlinc option at NSC, so it would be great if that could work everywhere."

R75:
"Prioritize convenience of users instead of convenience of system administrators"

R76:

"The only thing i can tell is that i had access to Lindgren at pdc (for a summer school i had there) and finding myself in a different environment translated in a significative slowdown of my work. I really prefer how Triolith, Kappa or Matter are working."

R82:

"The resources are very match my needs. And their operations are very good."

R83:

"Same job submission and software names and availability will help in shifting jobs between resources."

R86:

"I am working in computational astrophysics and am particularly interested in compact stellar objects (white dwarfs, neutron stars, black holes). My applications contain a large variety of physics ingredients (hydrodynamics, gravity, nuclear matter, neutrino emission, nuclear networks) and the corresponding codes have been developed over a time scale of ~ decade. They are (right now) all OpenMP-parallel and I am looking for computing resources in Sweden to run them on. (As an illustration: some of my larger simulations have been running for six months on 256 cores at the HLRN in Hannover, Germany.) So far, I am not aware of resources in Sweden where I could run comparable simulations. Therefore, I have applied again for computing time in Germany, which is of course rather cumbersome.This lack of larger shared-memory resources is the main reason why I have not made much use of SNIC resources."

R90:

"Large memory needed for high-level QM calculations.If such nodes were available, we would probably use it much more"

R91:

"For (some of) our programs, we need large amounts of memory per *core* (about 20 GB), and thus also fast and efficient cores (to not uselessly spend the money on more memory for many inefficient cores).We also need nodes with fast and large scratch disk, on the order of 1 GB/s, and 10-20 TB. Expressed differently, a reasonable disk speed would be ~200 MB/s per core.For this job, hyperthreading has ~0 impact, so such capable cores count as one core."

R99:

"Regarding capacity vs capability computing: in the field I represent (computational materials science), there is a great need of high capacity computing for medium-sized jobs MPI jobs. A high-throughput resource with

an island design (islands spanning 1 to a few Infiniband switches) would be very beneficial for handling large amounts of e.g. typical VASP jobs. Regarding common system configurations: the great hurdles for moving between center are not technical differences such as queue systems or login procedures, but rather access to scientific software, and difficulties with moving/access data that you have stored at another center.There is also a lack of software quality assurance, meaning that sometimes simulations give different results depending on which resource you are running on. This scares the users, and effectively kills mobility: if you start a compute project on one resource, you stay there."

R120:
"Geographical proximity of hardware is not important; however, proximity and availability of *people* (application and software experts) is!!!Memory per node is not always what matters; we often need large memory (few GB) per single-core job"

R125:
"My research would benefit most from having the darn filesystems gulo and pica stop hanging all the time. That should be priority #1 one, I don't care about any new clusters, or virtual machines or syntax harmonisation. Make the filesystem work first."

R127:
"Memory for us is a function of memory per CPU core on each node. We need 4-8GB of memory per CPU core. (I have no idea how to translate this into per-node memory since I don't know how many cores you are assuming per node.) Also please note that when you asked about the job type, we tend to submit 100s-1000s of single-threaded jobs all at once. I marked this in the intermediary categories because it wasn't clear if those meant "single-job, parallel" or "single-submission, many CPUs"."

R128:
"Very large scale calculations, more than 1000 cores, are normally for data production, but not for new science and new discovery. At least, we have not seen any exciting new results yet from those very large scale calculations. "

R133:
"Although I would highly benefit from similar procedures for all the above things, I think e.g. the job submission procedure at all clusters I have used is already sufficiently similar. However, the same software selection and naming would simplify things a lot, as would having a common login name for all systems."

R151:

"I am dissatisfied of Triolith (specially after the upgrade), the scheduling is horrible and I hardly get my jobs running (some 2 hr 8 nodes jobs starts after 16+ days and sometimes after 42+ days, so, I forget what that job was doing). Moreover, I believe (by 99%) that Triolith gives higher senior staff better priority, so, even they use much, they always get into the queue and fill the allocated core-hours.My second satisfaction is Abisko, which is histerically behaving, sometimes, you get your job running immediately, some other times, you get continous delays (for 10+ or more days, also for jobs with 8 or 4 nodes at 10 hours!). Also, Abisko can change the status of your job from priority to resources to valid or so several times and sometimes I never get my job running and cancel it myself. On abisko too, I believe for 99% that certain priority in the queue is given to senior researchers.The best out of them is Lindgren, as they have a clear quota system and I get my jobs running faster, the only drawback is: they allow small jobs to run, and that fill the allocated core-hours while lindgren is designed for large jobs (with 512 cores at least) so, I continously get blocked, but, I hacked that drawback, by splitting my large job of 24 hours into small time restart jobs of 2 hours or 4 hours each, so, sometimes I get my job running before the small jobs submitted by others. I feel so bad that good system like that designed for large jobs allowing small jobs to run..specially that I knew that this is so bad on the Lindgren system itself as it is designed for large jobs.For Zorn: it is quite good experimental system, but definitely, needs more RAM."

R153:

"The Lindgren cluster at PDC has been difficult to access for members of our team that use Windows as desktop platform. The Kerberos login-in procedure is considerably more problematic as compared to e.g. the ssh approach at NSC and HPC2N centers."

R157:

"Different compilations of software leading to different outcomes."

4 SNIC Storage Resources

Q44: *As a user of SNIC storage resources, what type of research data do you mainly store?*

This was a checkbox type of question where more than one checkbox could be ticked. This makes the total percentage figure higher than 100.

Data type	Ratio
Data that cannot be regenerated	23%
Data that cannot easily be regenerated	65%
Data that can easily be regenerated	33%
No answer / Not applicable	10%
Other	1%

Q45: *If you are a user of SweStore, how often do you copy/stage the data from SweStore to other resources?*

This was a non-required question, which is why the total percentage figure is lower than 100.

Data staging frequency	Ratio
I never copy/stage data.	6%
I occasionally copy/stage data.	17%
I frequently copy/stage data.	1%
No answer / Not applicable	65%

Q113 – Q115 are ratings based on the input request: *How do you rate the interfaces to access SweStore data.* The answers rate the alternatives on a scale with levels "Poor", "Less than satisfactory", "Satisfactory", "More than satisfactory" and "Excellent". The results are summarised in figure 8 as the ratio in percent of all responses. The questions were non-required and therefore the total percentage figure is lower than 100.

Q113: The SRM and GridFTP interfaces to dCache. **Q114:** iRODS i-commands. **Q115:** Web browser.

Q46: *Where is your research data stored?*

This was a checkbox type of question where more than one checkbox could be
ticked. This is why the total percentage figure is higher than 100.

Storage location	Ratio
SweStore	20%
SNIC centre storage where I do research	75%
On my laptop/workstation	52%
On group/department/university storage	43%
On equipment where experiments were conducted	9%
Other	4%

Q47: *Are there any legal or ethical issues to consider regarding parts of your research
data?*

Legal issues?	Ratio
Yes	9%
No	67%
I do not know/No Answer/Not Applicable	24%

Q49: *Have you needed a system for requesting storage allocations similar to the
SNAC CPU time allocation?*

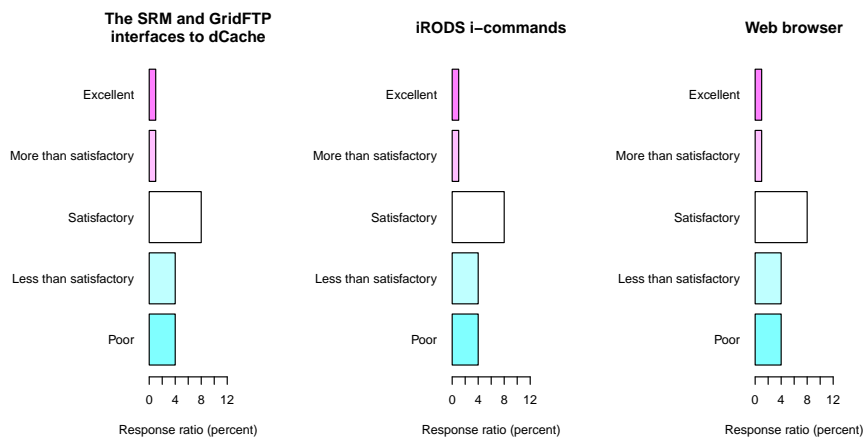| Needed storage request system	Ratio
Yes, for SweStore	13%
Yes, for centre storage	14%
No	77%

Figure 8: Respondent rating of the different interfaces to SweStore. The percentage figures refer to
the total number of respondents.

Q50: *Does your research group have a data management plan?*

Has management plan	Ratio
Yes	20%
No	61%
I do not know	19%

Q51: *Is the geographical proximity to SNIC storage hardware important to you?*

Geographical proximity is important?	Ratio
Yes. Please specify why on the next page.	5%
No	74%
I do not know	12%
No Answer/ Not Applicable	9%

Q52: *The geographical proximity to SNIC storage hardware is important to me because ...*

R32:
"Usually I must download all the data I have generated in the SNIC storage. How close I am to storage, the speed of this procedure increases.In fact, in my case, the connection to the SNIC storage is very low. Sometimes I have lost the connection while in access, and because of this I have to ask for downloading all the data again."

R44:
"I think I am missing the key point of this question. I need fast internet connection to retrieve my data. Fast internet connection may be related to geographical proximity. "

R52:
"As a system expert, I may need to inspect and handle the hardware."

R70:
"We transfer large amounts of data on a daily basis."

R118:
"Mainly because of the speed of transferring data"

R119:
"proximity to experts"

Q53: *What research data services do you consider to be most useful for your research and that you expect SNIC to deliver?*

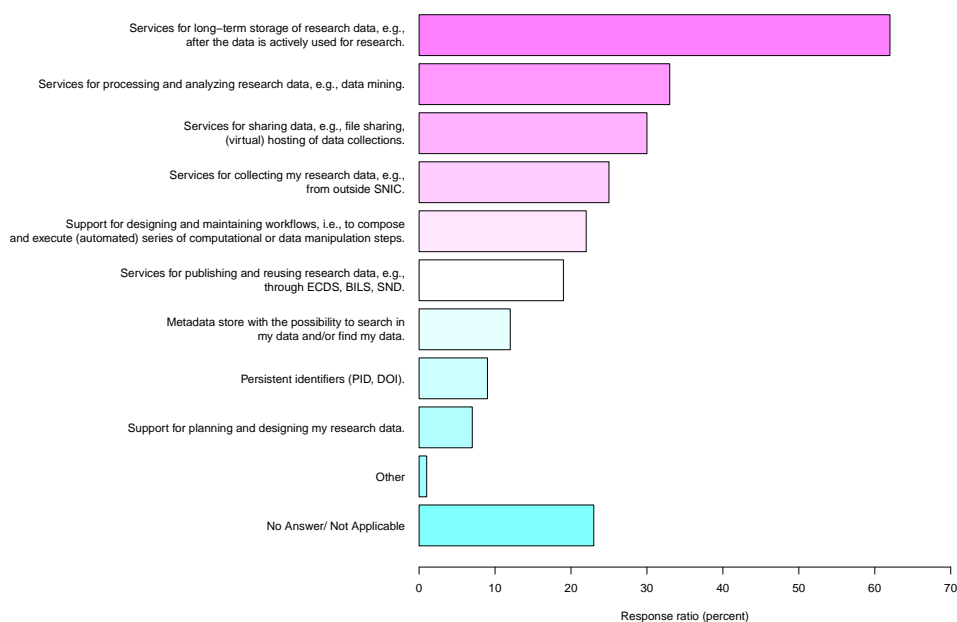[**The answers are summarised in figure 9.**]

Figure 9: The data services respondents wanted SNIC to provide.

Q54: *Please provide any further comments that you would like to add concerning SNIC storage and research data services. In particular in case you have expressed your dissatisfaction in any of the answers, we would be grateful you could provide more detail.*

R21:
"There's very little good information available about how to apply for and use SNIC storage. A set of guidelines for usage and methodology (ie. how to effectively name, tag, index, archive your data) would be greatly appreciated."

R25:
"THe currently handling of the data is good, despite the initial small quota. However i'm not sure how it will be with the implementation of the new storage at NSC, but i believe that will be ok also."

R34:
"machine learning & Bayesian statistics tools"

R37:
"The interface between SWESTORE and the local storage should be made significantly easier. I suppose the syntax is more or less simple, but it's still not standard and completely analogous to other folders and drives on the computers I have access to. The ideal would be to have SWESTORE as just

24

another folder on the file system, so that I could just rsync everything over, look at what files are stored and perhaps even do some simple data analysis or at least reading of text data directly in the folder. It would also be nice if it could function as a node in something like a dropbox/bittorrent sync system. That way I could have it as a part of a live backup system for the research data."

R39:
"As fairly new to MD simulations, I have done some data management work (like having local backups and long time storage), but I think it would be good to have some sort education about best practices of data management. -how to use various various SNIC resources- advise on setting up a local data management plan- education on best practises for various types of applications- a 'boot-camp' to send new postdoc and students to, to make it easier to maintain best practises."

R43:
"It is important that if large-scale data storage has been listed in a SNIC application, a center the project is allocated to can actually provide that storage."

R47:
"As mentioned I believe that it would be of great interest to support distributed databases as Cassandra and Hadoop, both in terms services to users but also to improve the overall competence on these databases in the universities, but also as a catalyst to Swedish industry."

R48:
"I would like to say that I appreciate the current SweStore storage system a lot. Having somewhere to store large amounts of data is useful."

R53:
"My group has only long-term storage plan, I think.Services for sharing data, e.g., file sharing, (virtual) hosting of data collections. - I think it very confusing, esp. for new students/colleagues, with data sharing between me (teacher, coworker) and a student/colleague. It would be excellent to have some DropBox type possibility to be integrated between SNIC users."

R59:
"We need computational power. Not data services."

R82:
"In the past I asked for more 1000 GBi volume to store my data. I very satisfied to get it but as far as I understood the big disadvantage is that this data is not backup."

R91:

"As a maintainer of some (~25 TB) local storage, I am very happy that SweStore exist. It makes it possible to direct users with larger requests to SweStore and more and more use the local storage as work-space."

R120:

"Like mentioned above, geographical proximity of hardware is not important; however, proximity and availability of *people* (application and software experts) is!!!"

R121:

"I am a beginner to using Swestore and do not fully understand all the questions here. We found that the procedure for getting a certificate was quite complicated and now I have questions about how to use it: I have the certificate installed on my Mac and when I go to the https://webdav.swestore.se/browser/ I access my files directly, without any additional log-in. I don't know how to access my files from other computers, if necessary, if that is even possible."

R124:

"Access to SweStore is always a mess. I make it work, then suddenly it stops working because I have to renew some certificate and then usually the whole process for handling certificates have changed from previous year and you have to go through a new process. Why can't you link your SNIC account to a SweStore allocation by default? Frankly, if you are not working a lot with certificates you have no idea of the different CA, VO, Terena, Nordugrid whatever you have to be part of.and as always when the SNIC "

R125:

"My research would benefit most from having the darn filesystems gulo and pica stop hanging all the time. That should be priority #1 one, I don't care about any new clusters, or virtual machines or syntax harmonisation. Make the filesystem work first."

R127:

"Easy ability to store data at different levels of redundancy. E.g., centralized but not backed up vs. backed up."

R142:

"I am very aware of the benefit - and we are starting up on that but we are behind. In time I will be needing serveices for designing workflows but I answered no because I am not there yet."

R153:

"Data storage at SNIC resources has historically been associated with individual users, not particular projects. Any transition towards project based

storage should keep this history in mind allowing for a considerable transition time so that the users can adapt to a new way of thinking about data storage, without to much interference in on-going research projects. Also, several users work across different projects. Such usage should not be hindered by storage issues."

R160:

"I did not know about iRODS at all. I generally find the middleware for SweStore somewhat archaic (arctools, lftp). The web interface is good though and possibly iRODS would be better, I would need to try it.The documentation for SweStore could do with an overhaul:
http://docs.snic.se/wiki/SweStore#Getting_access does not even mention iRODS."

5 SNIC Support Services

Q55: *How do you rate the e-mail based support that is provided by SNIC?*

Satisfaction level	Ratio
Poor. Please specify below.	0%
Less than satisfactory. Please specify below.	1%
Satisfactory	30%
More than satisfactory	30%
Excellent	31%
No opinion	9%

Q56: *Further comments regarding e-mail based SNIC user support. (Especially if you answered 'Poor' or 'Less than satisfactory' above)*

R13:
"Often fast and good, but once in a while the response can be a bit slow. "

R17:
"Both PDC & NSC have improved their user account generation time tremendously, and we now typically get them within 1-2 days. This is much appreciated."

R25:
"The support is good, especially because i'm a heavy user of it!"

R26:
"I have found user support to be very good, particularly from NSC."

R27:
"Fast replies! Very good."

R44:
"I am marking this because I am trying to average all the centers. Some centers are excellent. and some centers are less than satisfactory and arrogant. "

R64:
"Uppmax provides great support."

R75:
"I am quite pleased with the e-mail support provided by NSC - very professional, on time and pretty much always resolve my problems. Sometimes I'd like the response to be faster but I recognize that it is not always possible. I had a negative experience, however, when I had to wait for response for months and the response have not resolved my problems. Such cases

28

should be given particular attention. I want to emphasize that the negative experience is not with NSC."

R76:

"I have had many problems in this year and a half of my PhD and i always found the service excellent. Fast replies to e-mails and always trying to come back with a solution and try to understand the problem.For some particular issues, that were difficult to solve via e-mail, it became necessary a very helpful meeting with an application expert.In conclusion i have always been able to obtain support in terms of advices and/or solution of problems."

R82:

"I very satisfied from the way you handle problems I was encounterd. Thanks!"

R98:

"When the number of emails grows between USER and SUPPORT TEAM, it is not easy to trace what has happened. Alternatives like BUGZILLA or the like are preferable."

R99:

"(I am partially evalulating myself here!)The response times can sometimes be long. I believe this is not really due to a lack of resources, but rather lack of routines in the support system.In many cases the users could have gotten an initial reply much faster, asking either for additional necessary information to solve the problem, or a time estimate when we can start working on the case."

R121:

"I don't know enough about SNIC yet to answer all questions"

R127:

"Very rapid responses to my UPMAXX requests! Quite happy."

R134:

"It really depends on the person on the other side. I've had both great and not so good experiences, and it all came down to the name signing the email. There are black sheep in every profession and/or organization/company. However it would be unfair to give a lower grade to the overall support, because there are many good people there too, which have provided excellent support throughout the years."

R144:

"Possibly Im mixing up things here and I have troubles distinguishing what is SNIC and what is NSC. The user support at NSC is Excellent! I don't know about any other user support from SNIC..."

R152:

"Good response time. Helpful and competent service."

R155:

"Group members have always received swift and useful responses when asking questions. "

R157:

"Technical user support at Uppmax has been quite poor. Much better at NSC. Response to issues related with SNIC allocations has not been good either."

Q57: *How do you rate the online support end-user information that is provided by SNIC?*

Satisfaction rating	Ratio
Poor. Please specify below.	0%
Less than satisfactory. Please specify below.	4%
Satisfactory	50%
More than satisfactory	20%
Excellent	9%
No opinion	18%

Q58: *Further comments regarding end-user information provided by SNIC. (Especially if you answered 'Poor' or 'Less than satisfactory' above)*

R7:

"Quality and structure of system documentation varies significantly between centres."

R17:

"My impression is that there is lots of duplicated simple information at all the centers, but relatively little advanced information - that we typically find at foreign centers instead."

R53:

"Your guides on using your systems are crap and even have errors sometimes. The examples are very far from a real life usage. "

R55:

"Information on trouble with the computing resources is very poor. I lost many CPU hours and working hours due to problems with the system, which could have been avoided, if information was provided on the webpage, or even better by e-mail. From discussions with colleagues I am convinced that many users would be happy to get a message whenever there are problems with the computing resources they are using."

R58:

"Ideally the community/users should be involved in developing SNIC services for their own benefit. SNIC pages could potentially be complemented by linking to community generated pages or somehow involving the community when creating information on how the community could benefit from SNIC resources. Compare for instance SNIC pages:
http://docs.snic.se/wiki/Category:Structural_biology
https://www.nsc.liu.se/support/systems/triolith-getting-started/
with corresponding community generated pages:
http://psf.ki.se/Xray/PSF Xray Software 2013.html
http://psf.ki.se/Xray/NSC settings.html

Maybe SNIC pages could have a commentary field to get community input on services provided? or maybe the community could contribute directly to the SNIC pages?"

R63:

"It is rather uneven. Some things are thoroughly documented, some not. In some instances I've run into information that is seriously out of date or premature (i.e. published before it was made to work, and then never finished). That is probably unavoidable with a documentation this size, though."

R68:

"Some information about hardware, such as cache configuration of the nodes, is missing (or at least too hard for me to find).The instructions on how to submit jobs are good!"

R99:

"The information about, for example, available software, is incoherent, spread out in many places (center pages, SNICdocs, individual resource groups, software project sites etc), and not always up to date."

R124:

"There is often too much information. If the goal is to achieve something I only need one path to achieve that goal.The perfect example is creation of new user in nsc express. First there are 3 alternatives for login, then you have the click the link "If All Alternatives Above Fail", then you have read some more alternative, until finally you are presented at the text "if none of the above applies create a new user" and the actually link to create a new user.This is of course just one simple example, but there are many more, for instance getting access to SweStore, again multiple options, certificates solutions, converting certificates, you name it. "

R127:

"The web site is hard to use. I have to search for where to make a new user account every time I want to get a new PhD student up and running. "
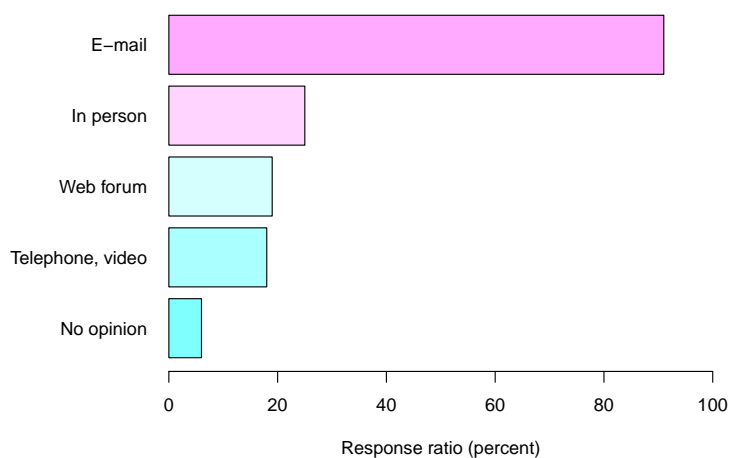
Figure 10: Preferred ways to interact with SNIC support

R141:

"More and uptodate compilation details would be valuable.And also that you make available somewher in the systemhow you compiled the software you install,so that I could recompile a later version or alter the code. CPMD and CP2K for example."

R160:

"See my comments above about SweStore documentation at SNIC. Much of the documentation provided at docs.snic.se is not very useful for me."

Q59: *In case you have access to SNIC resources at multiple SNIC centers, how do you perceive the quality of the support that is provided?*

Center support equivalence	Ratio
The centers provide the same (or similar) quality of support.	22%
The centers provide support that is of somewhat different quality.	19%
The centers provide support of widely varying quality.	9%
No opinion.	50%

Q60: *How would you prefer to interact with SNIC support?*

[**Responses are presented in figure 10. Multiple answers could be checked for this question.**]

Q62: *Are you familiar with the SNIC applications experts?*

Familiarity with Application Experts	Ratio
Yes	48%
No	52%

Q63: *Have you received help from a SNIC application expert in the past 24 months?*

Have received help	Ratio
Yes, I receive help from the SNIC application experts regularly.	9%
Yes, I receive help from the SNIC application experts occasionally.	23%
No. There has been no need for application expert help.	7%
No. There was a need for application expert help, but there is no expert in (or close to) my scientific area.	1%
No. There was a need for application expert help, but I did not know how to contact the SNIC application experts.	0%
I don't know	1%
No answer / Not applicable	4%
Other	2%

Q64: *How do you rate the support that you received from the SNIC application experts?*

Support rating	Ratio
Poor. Please specify below. [**See Q67**]	1%
Less than satisfactory. Please specify below. [**See Q67**]	1%
Satisfactory	6%
More than satisfactory	12%
Excellent	13%
No opinion	13%

Q65: *What type of support efforts do you think SNIC application experts should focus on in their work?*
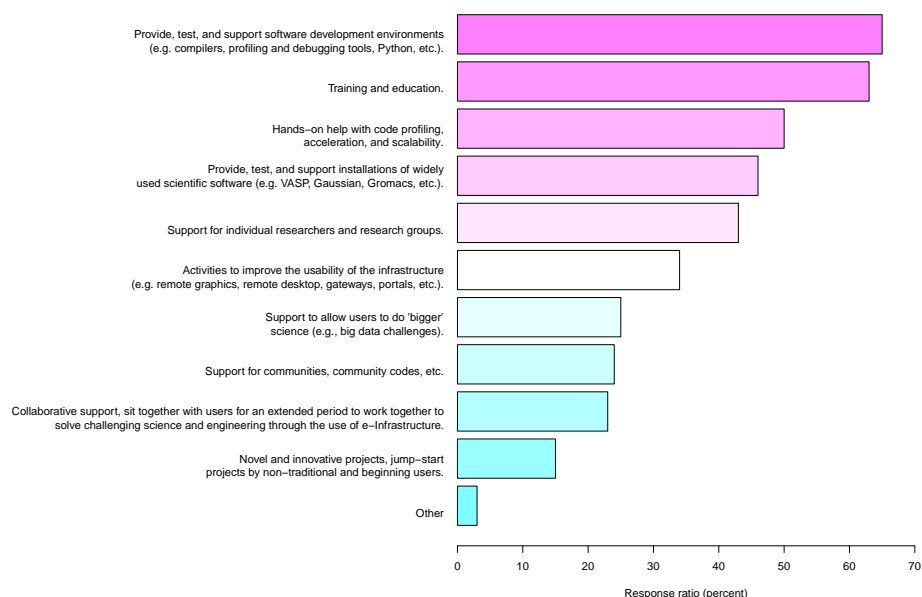
[**Responses are presented in figure 11.**]

Figure 11: Respondent preferred application expert focus areas.

Q66: *What type and size of efforts do you think SNIC application experts should undertake predominantly?*

Effort type	Ratio
Continuous efforts to build, develop, and maintain the application level environment on SNIC systems (e.g. software installations, development environments, etc.).	46%
Short-term support projects, up to 6 months.	9%
Long-term support projects, more than 6 months.	3%
Both long-term and short term projects.	17%
No opinion	25%
Other	0%

Q67: *Further comments regarding help from SNIC application experts. (Especially if you answered 'Poor' or 'Less than satisfactory' above)*

R17:

"We primarily use the advanced application experts that were funded as part of the SFO initiatives, but unfortunately we get the impression most of them have to split their time and do quite a lot of day-to-day helpdesk duties too, rather than focusing on their advanced expertise."

R25:

"The support is quite good. Having a application expert with the program

34

that i work is excellent. It decrease the amount of time expend compiling and debugging by "at least an order of magnitude".And his tips are good! Keep it running!"

R33:

"Application experts need to work closer with research groups. Occasional hour long or even one-time/rare few hour meetings can provide only little support for more complex issues which are often the more pressing ones (e.g. scalability of codes, best practices to use resources and codes efficiently, etc.)."

R75:

"I want to praise the help I receive from Peter Larsson at NSC. His extremely professional attitude and thorough knowledge helps me a lot,including my use of PRACE computers."

R99:

"I problem I see, is that I close many support cases in the RT, due to lack of further response/interaction from the users that initiated the request."

R124:

"The whole system with AE is not working. The SNIC centers have competence in computing but they have no competence in recruitment in the applied areas. There is a need for a much better integration of the AE in the research areas. Maybe there is a need for a new class of advanced support giving people work much closer to applied areas like Research Engineers."

R140:

"The expertise was not competent"

R157:

"Very good service form Peter Larsson at NSC. Poor service from previous Uppmax application expert(s)."

R159:

"I do not really understand the role of the application experts. We are experts in the methods and their use. We would instead need help to get the codes running as efficiently as possible on the different machines."

Q68: *How would you like to interact with SNIC application experts?*
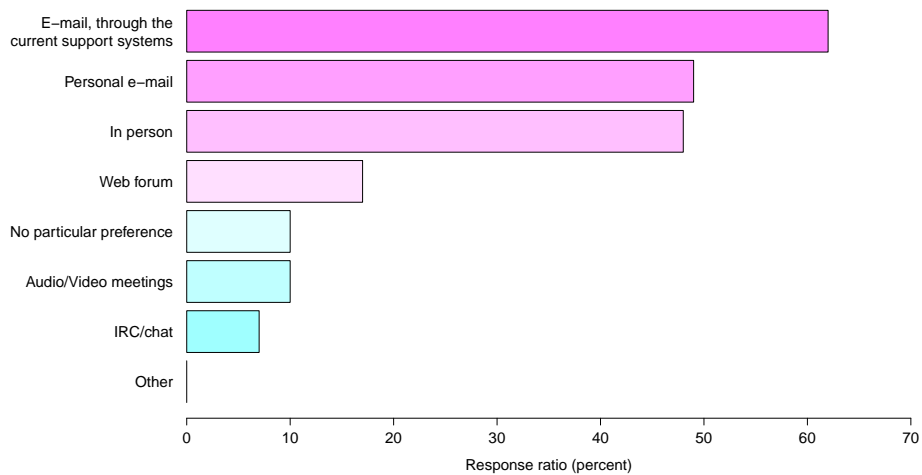
[**The answers are summarised in figure 12.**]

Figure 12: The respondents preferred way to interact with the SNIC application experts.

Q61: *Please provide any further comments that you would like to add concerning SNIC user support. In particular if you have expressed your dissatisfaction in any of the answers, we would be grateful if you could provide more detail.*

[**No answers were recorded for this question**]

Q69: *Please provide any further comments that you would like to add concerning all types of SNIC user support. In particular if you have expressed your dissatisfaction in any of the answers, we would be grateful if you could provide more detail.*

R17:
"I think SNIC needs to split the current 'application expert' position into (1) technical experts that do system administration, installation support, and helpdesk duties and (2) advanced application experts that really have deeper knowledge than the advanced users.Today, SNIC is much too focused on the simple helpdesk duties, IMHO."

R37:
"I think the work that Peter Larsson (NSC, http://www.nsc.liu.se/~pla/) presents on his blog is a very good example of things that application experts can do to spread knowledge about proper optimization of codes, and about differences between different codes. I have especially enjoyed his benchmarks of different codes."

R91:
"I work as a (local) application expert, so know that if one wants to give some kind of useful in-depth help / optimisation support, it takes a *lot*

36

of time. But I also get the feeling that many codes that are used are quite inefficient... "

R99:

"I have often advocated the 'Pyramid' picture of user support. There is, I believe, no conflict between what is described as 'continuous effort' above, and long-term projects. One requires another: without a good infrastructure to build on, it is impossible to have advanced user support.For example, what use is there have application experts do code-optimization and algorithm development in a research group, if there are not even working compilers and tools available at the HPC site?Or: to help users set up an advanced work flow, the software components have to be there, in the first place."

R120:

"Support system should be the first entry point (1st line of support); further questions can be redirected to a 2nd line (or even 3rd if such will exist)"

R127:

"The application support people are not useful to us as we develop our own software so they can't really help with that. However, we have needed them to help us get system support for some of the virtualization tools we used installed on the clusters."

R153:

"Very good experience in interacting with NSC application experts, like for VASP and Wien2k support. "

6 SNIC Resource Allocation

Q70: *How do you rate the information that is provided by SNIC about resource allocation (e.g., mailing list, calls for allocations, on-line information)?*

Rating	Ratio
Poor. Please specify on the next page [**See Q75**]	0%
Less than satisfactory. Please specify on the next page.[**See Q75**]	2%
Satisfactory	45%
More than satisfactory	20%
Excellent	8%
No opinion	25%

Q75: *Specify your dissatisfaction with the information that is provided by SNIC about resource allocation.*

R7:
"Could be more clearly described and easier to overview."

R39:
"I have found it quite difficult to understand the allocation policies, and correspondingly helpful to talk to people "in the loop", that I did not get connected to via SNIC. A few times I felt that I got sub-optimal allocations because I did not manage to communicate what I needed well enough, or because the application for is not flexible enough.It might be a good idea to try give new users and applicants some help in applying, to lower the learning curve somewhat."

R127:
"I don't really know whether I should apply for a large grant or just keep creating new small ones."

Q71: *How do you rate the resource allocation policies?*

Rating	Ratio
Poor. Please specify on the next page.[**See Q76**]	2%
Less than satisfactory. Please specify on the next page.[**See Q76**]	7%
Satisfactory	44%
More than satisfactory	12%
Excellent	2%
No opinion/ Do not know	32%

Q76: *Specify your dissatisfaction with the resource allocation policies.*

R10:
"There seems to be a prevalent opinion (in the SNAC panels?) that SMHI is a government agency that shouldn't compete with universities and research institutes, and some of our proposals were substantially cut, other completely rejected. However, the point of regarding SMHI as a government agency with sufficient funding is outdated, SMHI's research department participates in national and international projects under the same conditions as any university or research institute, and we therefore shouldn't get "punished" when applying for SNIC resources."

R20:
"There should be larger resources allocation (> 5000 CPU-Hrs/month) for non-PI researchers. Specially for those who prove that previous resources has been used properly and the research results has been published. "

R24:
"The fly-time for a batch script is ridiculously short especially for GPU jobs"

R42:
"I am working on molecular dynamics simulations of protein-ligand interactions. This is a fast growing area and computers can play more and more important roles in elucidating the mechanism of protein-ligand binding. The simulations often need a lot of computational resources, in particular for systems involving large membrane proteins. I applied two times the large resource allocation and my applications were denied. I think the review panel shoud be more familiar the area and give an appropriate estimation of the importance of the projects involved in the application, instead of just using an excuse to reject the application."

R59:
"The divison between medium and large allocation, in combination with the "mandatory reduction" of the granted time, makes it more or less impossible to get a full medium project or a project that is in the lower range of large.Not all kinds of simulations can use thousands of cores, but they can still have a large need of computer power for simulations that need to run on a medium number of cores for a very long time. And several such simulations may be needed to be run at the same time.The review comments are not enough to determine why a project was not granted time, which makes it questionable on which grounds the decision was taken."

R75:
"The procedure of allocation is quite closed and based on wrong criteria.The results of allocation are non transparent and not available for scrutiny by applicants. This leads in some cases to dissatisfaction and frustration."

R85:

"Big bosses are eating all the computation times"

R99:

"I have gotten reports from PIs of low-quality evaluations in the scientific part (i.e. where the referees where making clearly nonsense remarks) and also of inconsistencies (like two similar applications being rejected in one case, and being granted in another).It is my personal impression that the more high-level political guidelines of SNAC allocations are not well-described or well-understood by the applicants."

R124:

"If you have received funding from VR. I don't see the point of having to apply multiple times for computer allocation in order to do the calculations needed for the application you already got funded."

R133:

"We have short periods of high-demand usage and then long periods with no CPU use at all. Above that, we need extensive amounts of disk space and RAM. This combination is virtually not accounted for in the allocation policies."

R140:

"Should be coordinated with VR"

R144:

"I think that the process is backwards. The idea of first applying for grants and then to have to try to find resources once again writing a new proposal is strange to me. Somehow infrastructure should be included already in the first scientific applications.As the system works now getting financial support for a project is not a guarantee that the project can be completed which is highy unsatisfactory. Further, having access to computing resources is often a prerequisite in EU proposals."

R151:

"The small allocation is really small, where I do need it and applied for it in order to help me get my jobs running away from the common group project that experience large queuing time. Small allocation should be with more hours, or maybe you can solve that by dividing the medium and big allocations on the group users equally or set it up to the PI to decide on the percentage of different users, so, each one can have a fair share."

R154:

"For the small and medium size allocations it is very clear what information that is expected/wanted by SNIC. However, for the large scale allocations it is unclear what the balance should be between the different types of

information, e.g. background of the research issue, relevance, scalability,benefits, requested CPU hours. ”

R157:
“A few times a negative remark purely about the science in the proposal (but not even discussed in the proposal) has led to allocations being reduced by 50%. Once the "yearly" allocation is reduced it continues on the same level the next year, as it is merely a copying from one year to another. ”

R159:
“Jag förstår attd etta är svårt men: Jag har skrivit ansökningar till snic i kanske 10 år, men jag har aldrig fått någon feedback på ansökan mer än antalet allokerade timmar. Det vore bra om med någon kommentar har ansökan uppfattas och om man vill att något skall skrivaas/göras annorlunda nästa år. ”

Q72: *How do you rate the integrity and transparency of the resource allocation process?*

Rating	Ratio
Poor. Please specify on the next page.[**See Q77**]	3%
Less than satisfactory. Please specify on the next page.[**See Q77**]	7%
Satisfactory	43%
More than satisfactory	7%
Excellent	1%
No opinion	39%

Q77: *Specify your dissatisfaction with the integrity and transparency of the resource allocation process.*

R10:
“It is not clear on what grounds proposals are cut or rejected.”

R33:
“I got the impression that a simple maximized utilization rather than efficient utilization is the main (sole?) goal, and judging the nature and suitability of computational workload for the respective HPC resource is left for the review and likely never contested later. This way the incentive to actually ensure efficient use of resources is not a priority (e.g. don't scale beyond the limit of the code - just because one has compute hours, there is no point to run at 10% parallel efficiency instead of say 30% at half the node count).”

R42:
“Please see my comment before. ”

R59:

"See previous comment, regarding transparency. A declined application must be given detailed feedback."

R75:

"Please see my comment above. There is no transparency at all. As for integrity - I do not know. I do not reasons to doubt it, except that some decisions made contrary to a common sense."

R90:

"The transparency is low"

R93:

"The basis or criteria for evaluating the proposals are unclear to me. This could be since I'm fairly new user of SNIC resources. It has not been a problem for me though, as I two out of two times have been granted the CPU hours I requested "

R99:

"see previous answer"

R124:

"No idea how it is evaluated. Probability of allocation granted = rand() + 0,5(if VR funded)"

R125:

"One of my projects just went down from 50000 hours per month to 2000, but no email no warning. Why ?"

R127:

"I keep asking for more and more small grants and get them. Someone should be considering if this is appropriate and how much this is costing to give me. I don't see any of that so I have no incentive to not keep asking for more projects."

R130:

"no meaningful justifications are provided."

R151:

"- Queuing time is not reliable as I said before.- Small allocation is really small ...for PhD students who needs this small systems to get their jobs running avoiding the horrible queuing on the group project."

R157:

"Personally I failed to understand how a single person with a very modest scientific output could get a similar allocation on Triolith as a group of about 6 persons with a high-level scientific output. "

R159:

"Som jag skrev ovan, så är det svårt att förstå processen när den enda information som kommer tillbaka är antalet sökta timmar multiplicerat med en faktor på kanske 0.4-0.6."

Q73: *How do you rate the feedback from SNIC's resource allocation committee (SNAC) on applications, in particular regarding the (external) review and allocation?*

Rating	Ratio
Poor. Please specify on the next page.[**See Q78**]	3%
Less than satisfactory. Please specify on the next page.[**See Q78**]	4%
Satisfactory	28%
More than satisfactory	7%
Excellent	2%
No opinion	55%

Q78: *Specify your dissatisfaction with the feedback from SNAC regarding external review and allocation.*

R10:

"It is not clear why proposals get cut or rejected."

R17:

"My impression of the allocation system is that there is relatively little prioritization based on the scientific value of the results compared to the resources requested, but it is at least partly a matter of scaling down all applications. It is a bit strange that one of the most important criteria for getting a large allocation is the amount of time consumed in the past."

R34:

"As the science is already judged by funding agencies and their experts it sometimes strange to read read comments from people who either did not read the application or did not understand it.My (rather big) group was two years ago cut off all resources because my application was too long. I was thought a lesson I suppose but my group had to suffer. I wrote to the director to complain but he did bother to unswer. In the last round I was given my own application to evaluate (as a snac evaluator). I think it is a good idea as it gives a moment to reflect what you are doing. But overall I am very happy with snic resources and I understand that we are very priviledged in Sweden to have these these excellent resources (hardware, software and after all excellent people behind them) "

R43:

"For my recent large allocation, it turned out that one of the allocated cen-

ters could not provide the needed storage, even though this was written in the application as a bottleneck in terms of RAM and an important concern regarding disk space.It seemed as if the evaluators had not checked with the centers that our analysis of computer needs, and match to centers, was actually valid."

R59:
"See previous comments."

R75:
"It took me about 25 letters and 4 month to get a feedback on my application. I believe this is an extraordinary poor performance."

R124:
"Only for show, since they lack the competence in the applied area. It is pretty much useless."

R130:
"I never received any meaningful comments."

R141:
"A little bit more detail and transparency would not hurt"

R144:
"I have only limited experience of such feedback but the combination of a very strong reduction in the time applied for in combination with a more or less absent motivation was upsetting."

R157:
"The feedback was occasionally very poor, as for example it focussed on a single line in the proposal. The review should not be based on specific and personal views on a purely technical issue, but rather consider if the computer time is well and effective used in terms of scientific results and output. "

R159:
"Jag har redan beskrivit detta. Jag förstår att det är svårt men det räcker förmodligen med ett par meningar hur ansökan uppfattades. Utvärdering av vetenskapen görs ju på annat håll i det att de sökande har anslag från bidragsgivare såsom VR."

Q74: *Access for medium-size or large-size computing allocations is for 6 or 12 month periods. Does this match your research projects?*

Matches projects	Ratio
Yes	40%
No, I would prefer to have allocations with longer duration.	41%
No answer/ Not applicable	17%
Other	2%

Q79: *Computing allocations are constant over the allocation period in terms of core-hours/month. Does this match your workloads?*

Matches workloads	Ratio
Yes	57%
No. Please specify how you would like it on the next page.	23%
No answer/ Not applicable	19%

Q80: *How would you like your compute allocation spread over the allocation period?*

R2:
  "much better if corehours/year (or half year) instead."

R10:
  "Workload can vary during the project, there could be a preparatory phase to get everything in Place, then a heavy production phase, and finally a post-production phase. Preparatory and post-production phase are generally less demanding on the computing power."

R14:
  "It would be good if the allocation per months would at least partly vary e.g. by having a higher allocations if considerable less was simulated before (in addition to the short term priority changes)."

R17:
  "While the system is not a perfect match for us, I don't see any realistic alternative. Since the available resources are constant, the fair-share system in place at most centers is the best solution we have - no system will allow all users to save time to have it available whenever they need.However, in particular for the largest allocations I think it would be useful to have a mechanism where we apply for a fixed amount of total CPU hours for a specific project, to run e.g. during 3 months. This would encourage all of us to prioritize those projects - and be specific - rather than constantly soliciting SNIC for ever-larger monthly allocations."

R21:
  "Longer allocation periods would benefit my research as it takes a lot of

time to prepare and analyse data from my large systems. When running simulations I use up my allocation in one or two days. If I need to rerun or get additional data I get stuck in the queue for a long time."

R26:

"My compute usage is mostly testing the performance and scalability of new Gromacs implementations, which completes on irregular timetables and benefits from short turn-around times on short jobs on a variety of node counts (e.g. for producing scaling plots, or debugging performance issues). Obviously, that's not a good fit for most job-scheduling arrangements. Being able to spend compute allocation when I need to is essential - spreading it out 25% per quarter is infeasible because some quarters will have no significant need for use."

R27:

"I do not use the computing resource every month, but some months much more intensively. It is difficult to spread it over the months evenly."

R30:

"Not by month. Over the whole period."

R33:

"Molecular dynamics projects often have fluctuating amount of computational workload with moderate amount at the setup phase, peaking at the middle/production phase and often low compute needs during analysis. Hence, if is often the case that during times of production simulation, the uniformly distributed allocation results is resource starvation of the project."

R39:

"My workload typically varies slower than just over the course of a month, as I cycle between development, production, and writing. Averaging over 2-3 months might be a better model for me. In that case though, I think it would be useful to be able to set quotas (dynamically) for individual users within a project to have safe-guards against individual users gobbling up all time in multi-user projects."

R41:

"To be able to have a dynamic allocation (within certain limits) to allow for less use some months and more extensive use other months"

R43:

"We needed more in the beginning of the project, and would have benefitted from more flexibility."

R50:

"Ideally the user could specify the monthly allocation however they like

(up to the total allocation for the period, of course!)Also, unused core hours from previous months should be transferable to the current or future months."

R52:

"It should be possible to ask for a short-time increase."

R53:

"More like a fairshare scheduling - those, who had low usage one month could get better priorities next month, so that they can gain. "

R55:

"I have a strongly fluctuating need for CPU hours. When computing, my analyses are often CPU-costly, and I quickly am limited by the CPU allocation, while in other long periods I don't use any of the allocated time. A flexible allocation would be optimal."

R56:

"The workload vaires a lot at different stages of the projects, and a more dynamica allocation procedure might help. But I am not sure how that should be implemented."

R59:

"At the moment we use all time we are granted, because we have been granted much less than we need. If we would have been granted the time we need, we would of course have periods with more need an periods with less needs. Due to our current situation, I do not have an answer to the situation where we get what we need."

R62:

"Most often most resources are needed after a small testing period, but before additional complementary analyses are performed. Usage is usually not linear, but I understand if the easiest way is to allocate resources uniformly over the allocation period."

R68:

"I don't run jobs every month, but when I do, I need muuuch more than 2000 hours at a time. (This summer, I used >30 000 core hours during a month.) Would it be possible to ""save"" core hours from month to month? Allocation per user instead of project would also be better."

R71:

"For me it would be better to have a larger allocation eg the first 2-3 months and then have it lowered, since most of the compute-heavy things are done immediately (or quite soon) after the data arrives."

R75:

"Every researcher has periods of intensive computing and periods when the obtained data is summarized and assessed. Naturally, theuse of HPC resources varies accordingly. There are must be procedures that should allow for such variation. One, for example, would be a possibility to trade available computer time with other users."

R76:

"I find it very unfair that the projects have a global allocation for all the users added, there should be also some personal counting of the hours used, since many times i find all the project hours consumed by only one or two persons and everybody needs to wait days to run jobs"

R78:

"The system should allow for high use followed by low use on a time scale of a few months. "

R83:

"Maybe the unused hours per month should be shifted to next months quota! Some projects are computationally heavy at start and nothing later while others are maybe computationally heavy throughout or only at the end. So unused hours should be stacked! "

R89:

"The workload varies a lot more "

R91:

"The workload is not constant, so in that sense the allocations do not match. On the other hand, I cannot think of a better way to do it.It was a real improvement when it changed from monthly fixed to moving window monthly."

R98:

"Have a minimum core-hour/month and being able to increase it in certain periods (close to publishing a paper or participating in a conference or workshop) . We are ready also to reduce the allocation during some periods."

R127:

"It would work better to have a total number that can be spent over the time period rather than an auto-refill setup."

R130:

"It should be possible to spend resources earlier in the the year. Especially with > 10,000 core jobs, the monthly quota can be used up in a day or two. If this leads to demonstratable success (arXiv/preprint submission that

can be reviewed), the allocation of additional resources spread over the remaining period should be possible."

R133:
"I would like to have a total quota for CPU per year, as our needs vary a lot between months. Sometimes we use no CPU time at all, while at other times we need much more than our allocation in one month, while also requiring access to high-memory machines (>=128 GB RAM)"

R139:
"I now have 2000/hour month on UPPMAX (although more can be submitted at a very low priority). For my needs more flexiblity would be useful i.e ability to ""save up"" hours when not utilized. Alternatively getting 24000hours/year would suit me better."

R140:
"Our work is divided in development and usage so need for resoures varies a lot."

R143:
"Since we have a large difference of the size of the jobs and sometimes more intensive periods it is difficult to use the resource in an optimal way since priority goes down for a longer period when using more cpu time than allocated on a monthly basis. "

R144:
"Allocation periods should be much longer. Typically cycles (climate modelling) are of the order of 5-7 years. During such a period there is a model development and tuning phase when computational needs are lower and increasing. Then there is typically a very heavy production phase of 1-2-3 years when maximum allocation time is needed. "

R148:
"No idea, but there are periods when a lot of calculations have to be done, and sometimes there are interruptions with calculations for different reasons."

R154:
"If a situation occurs (due to debugging, illness etc) that we can foresee that we will not be able to use the CPU time for one month, it would be preferable to e.g. put the allocation on pause and shift the allocation one month ahead."

R160:
"The fixed number per month is too rigid. I understand that the centres want to spread the load evenly but there must be some flexibility in the

use of an allocation. Possibly one could work with a fixed number for 3 months."

Q81: *Please provide any further comments that you would like to add concerning SNIC resource allocation.*

R9:
"Would like to see support for GUI based tools, e.g, matlab. Would like to debug and run code GUI based, sometimes its easier to handle parallel runs with a GUI, not everyone is proficient with scripts and batch runs.Also GUI should be supported by all servers, that would be nice, since my project does not have access to all servers. "

R13:
"Generally good with a constant allocation level, but there could be improved possibilities to grant special allowances for particular needs that arise on short notice, such as running of jobs that require extra Resources (memory/nodes/total runtime)."

R24:
"GPU clusters time allocation should differ from homogeneous clusters time allocation as the GPU resources partition is much less flexible than the CPU resources."

R26:
"For my work, access to capability machines is required, but if those machines are filled with single-node jobs then not only is their network capability wasted, but the waiting time before a job can start is relatively high. I think there is a clear case for SNIC to target low-node-count jobs at particular resources (e.g. the previous capability machine), and either penalize the scheduling of or prohibit the run of low-node-count jobs on the newest capability machine.Supporting the suspension of compute jobs can be very beneficial in this area, and I have known it to work well at http://nf.nci.org.au/. The ability to suspend jobs means that large jobs that run for short periods of time can access the machine without leaving chunks of it idle while waiting for smaller jobs to finish.Further, a job-suspension policy that guaranteed maximum job suspension of (say) 25-50% of its requested run time provides encouragement to users to make accurate requests (rather than leave it at the maximum for the node count being used), which must improve scheduling and thus their turn-around time (whether they get suspended or not)."

R50:

"Ideally, the maximum size of a 'small' allocation would be a little larger. Perhaps 8000 or 10000 core hours per month."

R58:

"Moving large amounts of data (~1TB) into SweStore is often limited by data transfer capacity rate in Sweden or in Europe which is unfortunate however will hopefully improve in the future.Data transfer take so much time that sometimes the process is interrupted and has to be repeated. It is to my knowledge not possible to restart data transfer efficiently since searching through what data has already been transferred take almost as much time as transfer it again.Repeated data transfer due to interruptions are a huge time sink and waste of network capability resources. An algorithm for efficient restart of data transfer is needed."

R59:

"The queueing system priority at C3SE was previously based on calendar month. Not it is the usage the last 30 days. Setting priority according to the last 30 days gives priority to users who have been granted too much time, or who do sometimes not need as much time as they have been granted. We need more time than we have been granted, and we use all time we have been granted. Therefore we are punished by a low priority continously. For us it would at least give us better priorities every now and then if it was based on calendar month rather than the usage the last 30 days."

R75:

"Make the procedure of allocation transparent and public with all the relevant data available for scrutiny and easely accessible."

R82:

"I was VERY impressed by the fact that all the jobs I submitted (using VASP code) were not crashed or something like that. I also VERY impressed that I asked for 1000 GBi volume for storage and I got it without any delay. Thansk!Last, I think that the data on "nobackup" should be backup."

R86:

"As detailed above, for my current codes there is hardly any SNIC infrastructure (at least none that I am aware of). "

R91:

"Most resources are quite similar: many nodes with lots of (semi-slow) cores, little memory per core and little or small amount of fast (local) scratch space. Some diversity towards the other kind would be if interest."

R93:

"The two-step security access to log into LUNARC has giving me some

problems with transferring large amounts of data (~5 TB) from computing centres abroad (from colleagues I understand that I am not alone with this problem). I realize that security is important, but having to implement person hour intensive workarounds was not the best solution. An option for automated external access would be much appreciated. "

R98:

"There is an ongoing stream of new PhD students and other researchers joining and leaving a project. IF there is a central information place, like wiki, the new comers can be guided to learn what they need to access to SNIC resources. Otherwise, it takes along time to briefing people working in a project. This information center, wiki, may contains introductory pages on UNIX shell commands up to a more specialized commands for using a specific tool. For sure, you can have many more helpful ideas for this problem. I told just one."

R99:

"It seems to me that many of the big, established research groups, always have large and/or medium SNAC-projects running, as they are continuously being renewed.It would cut down on the administration and simplify planning for them if the allocation period were longer, or that they were simply given a certain "base allocation".The system with Large and Medium allocations does not seem to work as intended when there are research groups/department that have both several large and medium projects. My impression is that they are forced to apply for extra time through Medium projects, when they don't get enough computer time granted in the Large allocation."

R105:

"Sort of strange questions for employees at centra."

R120

"I have actually a conflict of interests here as an occasional SNAC evaluator, so I refrained from several answers"

R125:

"The allocation system doesn't work, everyone is part of many projects and when you have a job to run you just pick the one with hours still available. You can't really separate your research in projects like that.Same for when the quota of one project is full, you just use space in the one that has the most free space, independent of what the data is."

R127:

"Overall I'm very happy. But a lot of that is due to the fact that it appears to be a "free" service. I would honestly prefer it if we saw some of the cost

accounting so we could make more reasonable decisions about how to use this resource."

R151:
"more core-hours!forcing the users to obey the cluster rules (as on lindgren with large job submission, I think that can be done through suppressing any bash script with less that 512 cores.)"

R154:
"A continuous improvement and raise in the computer power is a necessity for researchers in Sweden to keep up with the competition worldwide."

R157:
"As written before, it seems that the way allocations are given is not transparent and in some cases it can actually be questioned on which grounds a distribution was made. "

R158:
"Since we rely heavily on SNIC resources for all of our research, it would be far better if longer duration could be provided. For example 3-year allocation, with a compulsory annual activity report and possible "upgrade" (also on annual basis) if deemed necessary by applicant and found reasonable by allocation committee. Having to apply every single year for projects that we know runs over several years it simply just time consuming - should better follow e.g. VR grant allocation or similar. "

7 General

Q82: *What should SNIC do differently? How can SNIC improve? Are there specific areas where SNIC can do better?*

R34:
"Maybe to increase compentence in GPU computing"

R38:
"On Triolith there is a ThinLinc service for remote desktop and visualization, which in my opinion is very good. It is my hope that the other SNIC clusters could offer similar services for remote visualization."

R41:
"Allocate the funding more evenly to the six SNIC-centra instead of having three very large and three minor SNIC-centra. The majority ocf the SNIC-users would benefit from more Medium-sized allocations and user support close to the research groups."

R52:
"I believe that it is important to have the SNIC centre in the geographical neighbourhood for the small and medium size projects, because this gives a local support and an easier step into the SNIC world. (When the projects have outgrown the medium size, they are more likely to be able to run at a remote centre, even if there is no guarantee for it.) SNIC has given too much money to the large centres and forgotten the needs of small centres. Here in Uppsala we have a large percentage of SNIC users, but not enough compute power to give them the medum projects that many of them need. The demand is already much higher than what Tintin can provide."

R53:
"Would be nice to have an effort on data communication with users: the current scenario is, I think, that I(=user) do a calculation on SNIC and get a result (=data), but the ways to access this data are not so developed. Basically there is only ssh as a ground level, and then there are different levels on top of ssh (e.g. sftp with some graphical clients for different platforms). I personally use high level (fuse) to mount my files as if they were on my laptop, but it is not very transparent and efficient. There was an effort on afs sharing from PDC, but it does not work through SNIC as different centers are in different kerberos realms, so one cannot simultaneously connect to two of them and switching is difficult. I have quite extensive experience from a more 'homogeneous' systems, like e.g. Mac to Mac sharing with afp or SUNSolaris, where the problem of sharing files between workstations simply did not existed. (UPPMAX had this type of sharing with UU computer network long time ago) I understand, that the system now is not so

54

homogeneous, but may be SNIC should make an effort either on software (like fuse) or hardware (adding some dedicated file servers to a cluster)? "

R57:
"Environment module defaults are a source of frustration - often new versions of software are available but the default is a very old version. This leads to me frequently having to re-run analyses after not manually specifying version numbers."

R59:
"For us the multi-core architecture is not ideal. Do not blindly count the number of cores, but also the rest of the hardware. The memory bandwidth from the cores is important as well. There are many kinds of scientific simulations that need more computer power, not only those that use thousands of cores to run many small problems. Are the clusters there for the users, or are the users there for the clusters?"

R62:
"SNIC should probably take the lead in how to work with genetic sequence data in a secure way (in accordance with "PUL" etc). It would also be great to have a national solution on how to store whole genome data, for instance at Swestore."

R75:
"Allocation procdeur should be changed."

R82:
"Backup the data on "nobackup" "

R86:
"Provide resources for larger, purely OpenMP applications. Already running 128-core OpenMP applications would be very helpful. Or, say, a large number of Ellen-type systems (PDC; or larger) "

R93:
"Within the climate modelling community, especially people dealing with past climate (geology, physical geography, i.e. general earth system science at universities) there seem to be a lack of knowledge about SNIC and what possibilities offered with the systems. Direct contact to these departments could be beneficiary. "

R98:
"Training users."

R103:
"No opinion"

R105:

"A better investment plan, so that new resources arrive more evenly."

R108:

"I would have more use for more Triolith like systems (good CPUs without a minimum limit on the number of cores to use)."

R120:

"SNIC still does not appear to be a *unified* e-infrastructure - rather, a set of disparate resources with quite different support levels. It is also rather odd that large universities, like Lund and Uppsala, have very modest resources and service levels to offer. While of course hardware might be anywhere, experts must be easily and promptly available locally, especially if one wants to set up a new software environment and such."

R124:

"More integration between the SNIC centers. I think SNIC would benefit from fewer centers, since it there is a base cost of maintaining a centre. "

R131:

"More training courses, both basic level and advanced level."

R133:

"More resources for bioinformatic sequence analysis, i.e. machines with lots of RAM but not large amounts of CPU cores. Access to large storage arrays for mid-term storage needs (1-2 years)."

R138:

"The time in queue is difficult to estimate, and the –start option often doesn't provide much information. I don't know if this is a local issue (in my case: for NSC resources), but it would be much appreciated if there existed some better tools for estimating start time.Small nodes easily accessible for development and small tests seems to be disappearing, and this is something I hope can be kept."

R140:

"Coordinate application with VR (and other infrastructures)Possibility for more variation between monh of usage,"

R146:

"SNIC should concentrate their efforts on the hardware side, both in terms of system and user training how to use the systems efficiently. On the other hand, on the application side, it is advisable that the e-Science centres (SeRC and eSSENCE) takes care of the organization and administration of application experts. Reduce number of computer centers to support and focus the resources on the the "big-three" (NSC,PDC and HPC2N)."

R153:

"Although generally good, an even more increase in focus on the need of the user, rather then focus on maximally optimal usage of each resource should be a priority. If the two collide, assistance to the user can be given, but if it is still not possible to combine, the need of the user should have priority."

R155:

"To me the most important points are 1) Good support to users 2) Education for users 3) Size of allocations. I find the support and education to be excellent. Of course the size of allocations is a matter of available resources. "

R157:

"Better / clearer way of deciding about allocations. Making more computing power and time available."

Q83: *How does SNIC compare to other national eInfrastructures (abroad) that you have experience with? What do other centers provide that SNIC should provide as well? Please list the center(s) that you use in the comparison.*

R12:

"I have experience of ICHEC in Ireland. They chipped in for funding of computational conferences / workshops, which was a nice gesture. And they gave away decommissioned hardware to local computer centers that wanted to keep using / upgrade such. Interesting practice. They also used international (no Irish) evaluators for project proposals, in order to not bias the established groups vs the new groups using HPC. Something that can be learnt from."

R17:

"I really like how XSEDE has transformed the US system to be entirely focused on the user experience and service rather than centers. However, an important part of that might be that they simultaneously created INCITE for the very largest projects. "

R22:

"Compared to the Galician Supercomputing Centre (Spain), the computing nodes of NSC are much faster, as well as the scalability that can be achieved here. However, storage did not have any quota limit in the former Supercomputing Centre, where it was also backed up."

R26:

"Compares well with the Australian national eInfrastructure - some similar problems with heterogeneity."

R33:

"More incentive and more resources for development on/for modern architectures. for instance, heterogeneous architectures are badly underrepresented and what exists is often outdated, hard to access, or ill-maintained. This is counterproductive as it avoids the chicken-egg problem of nobody will use the architecture because it is to new and nobody will adopt it because it's less effort to not do so - especially if such machines are not even available. It would be much better to incentivize adoption and research of new architectures by providing, next to the main HPC resource with the same file system, queue system a devel tools a smaller machine with "next-gen" architecture."

R34:

"Maybe to have a look at CSC in Finland which has a very broad profile. However, the service is not always as good as in Sweden"

R36:

"I have previously used TeraGrid in the US. SNIC is definitely better. "

R52:

"The small projects seems to be missing in Denmark. It seems like Gardar is the only opening for Danish users to get something like what SNIC calls small projects. This information comes from Lyngby."

R75:

"SNIC is very good in user support. SNIC is less good in providing hardware and amount of computer time."

R82:

"Not relevant."

R86:

"see aboveI am using and have applied again for resources at the Hoechstleistungsrechenzentrum Nord (Hannover, Berlin) in GermanyFor the research productivity of my group it would be very beneficialif our calculation could be performed on SNIC resources "

R88:

"Very good compared to Abel (good machine but poor service to get login information, certificates and file transfer)"

R91:

"I see that centre-storage solutions have become better lately. Please continue that. Lustre-style is needed, such that there is aggregate > 10 Gbit/s to the filesystems, and note that some few nodes completely can bottleneck the storage systems."

R103:

"Comparison with IDRIS - CNRS French center:As a simple users, SNIC works, imho and within my research activity, much better. SNIC system provide more freedom to users, less restrictions which could have prevented an optimized use of ressources. Let's say that a good compromise exists in SNIC. This results in a significantly better efficiency. "

R112:

"Fewer big computers."

R117:

"I used Archer in the past an it seems there there was a tendency to have more educational resources (webinars and frequent classes) to educate the users. "

R120:

"Through WLCG we have access to dozens of national e-infrastructures worldwide, and WLCG is an e-infrastructure by itself, albeit tailored for a small number of applications. In general, I have a feeling that many other infrastructures strive for more commonality in terms of interfaces and service levels."

R124:

"Well, I had computer time on NERSC this summer and I am actually much more happy with NSC."

R125:

"When I was working in Switzerland, the filesystem didn't hang like that. The queue was always free too."

R142:

"It fares well. Except perhaps in scale."

R146:

"Have used computers in Julich, Germany (the Blue Gene) in the past and SNIC still has some way before it reaches the same kind of close and swiftly support, but it not far behind. "

R152:

"SNIC compares well to other centers."

R160:

"I think that there are too many SNIC centres for a country as Sweden. It seems that every big university needs to have its own supercomputer centre but I do not see the need for that from a national perspective. I realize that this is a system that has grown over the years and may bedifficult to change but I suspect that fewer centres would make it easierto manage SNIC and

SNIC resources as well as allow larger systems tobe acquired.It may be good to compare the landscape in Sweden to that in other countries in Europe. I have not done this but I think that you will find that the numbers of centres per capita is rather high.”

Q84: *Do you have a need for eInfrastructure services (now or in the next 24 months) that are currently not provided by SNIC?*

R26:
“No”

R34:
“I would need competence in analyzing very irregular datastructures for pattern recognition and artificial intelligence and machine learning of not complete data.”

R39:
“For my application (MD), I would probably be better served by a focus on raw capacity rather than other types of resources.”

R52:
“As already said, larger compute resources for at least small and medium projects at UPPMAX.”

R53:
“File sharing? ”

R59:
“NO!”

R88:
“No”

R90:
“Large-memory nodes”

R91:
“Ahh, see answer on previous page.Nodes with:- fast *efficient* cores- > 20 GB/core memory- fast and large disk. 200 MB/s / core, 10-20 TB”

R98:
“Visualizations and profiling parallel application in scale of more than 20 nodes and 200 cores.”

R112:
“I see a need for large-scale facilities.”

R117:
"Larger fat node resources, similar to ellen at PDC but larger size for the short term use. The uses of these resources are usually limited to reduced size models for data reduction in postprocessing stage and the outcomes are usually outstanding. There is not a clear policy at the moment for the users how to use these (something like a queue system). "

R120:
"Transparent (common) accounting and (common) monitoring and information services are really really missing - not as static tables on a Web site, but real-time status information and historic accounting/usage statistics, overall and per project."

R145:
"No - HPC is the most important and should be the focus"

R146:
"Possibly need for visualization services"

R155:
"No"

R159:
"No"

Q85 – Q106 asks for respondent preference ratings of future focus areas for SNIC with the question "*What areas should SNIC focus on in the coming years? Use the ratings such that the areas that you consider least and most important are highlighted*". The rating levels are "No opinion" and 1 – 5, where 5 indicates highest preference. The results are presented in figures 13 and 14. There is a slight reordering of the rankings between the graphs corresponding to the difference between sorting on the mean or the median of the data within each category. The reason for including both is that neither is the obviously correct way to present this data. Overall they paint a very similar picture though.

The 'No opinion' option was in both figures taken to mean zero. All user answers were normalised to one in both , i.e. one awarded point got the weight 1/(total points awarded) on a 'per user' basis. The vertical black lines in figure 13 indicate quantiles 25%, 50% and 75% from left to right. The notched box plot in figure 14 follows standard conventions for this type of plot.

Q85: *Compute hardware investments, increase capacities,* **Q86:** *Compute hardware investments, increase diversity of hardware,* **Q87:** *Storage hardware investments, increase capacities,* **Q88:** *Storage hardware investments, increase diversity of hardware,* **Q89:** *Quality of operations (e.g., availability and reliability*

of resources and software), **Q90:** *Quality of user/help-desk support,* **Q91:** *More user/help-desk support,* **Q92:** *Quality of the online support information and end-user documentation,* **Q93:** *More online support information and end-user documentation,* **Q94:** *Quality of application support,* **Q95:** *More application support,* **Q96:** *Quality of the resource allocation process,* **Q97:** *Provide better mechanisms for resource allocation,* **Q98:** *Service description, i.e., document what computing and research data services SNIC delivers,* **Q99:** *Harmonization of user interfaces across similar SNIC resources,* **Q100:** *Information to end-users (e.g., allocations and news),* **Q101:** *Interaction with users to identify their needs and satisfaction,* **Q102:** *Support the life cycle of research data, from the moment it is created to when it can be deleted,* **Q103:** *Visualization tools support,* **Q104:** *Education/training of end users,* **Q105:** *Reach out to potential/new users,* **Q106:** *Projects to explore emerging standards and technologies, e.g., clouds, possibly in collaboration with users,*
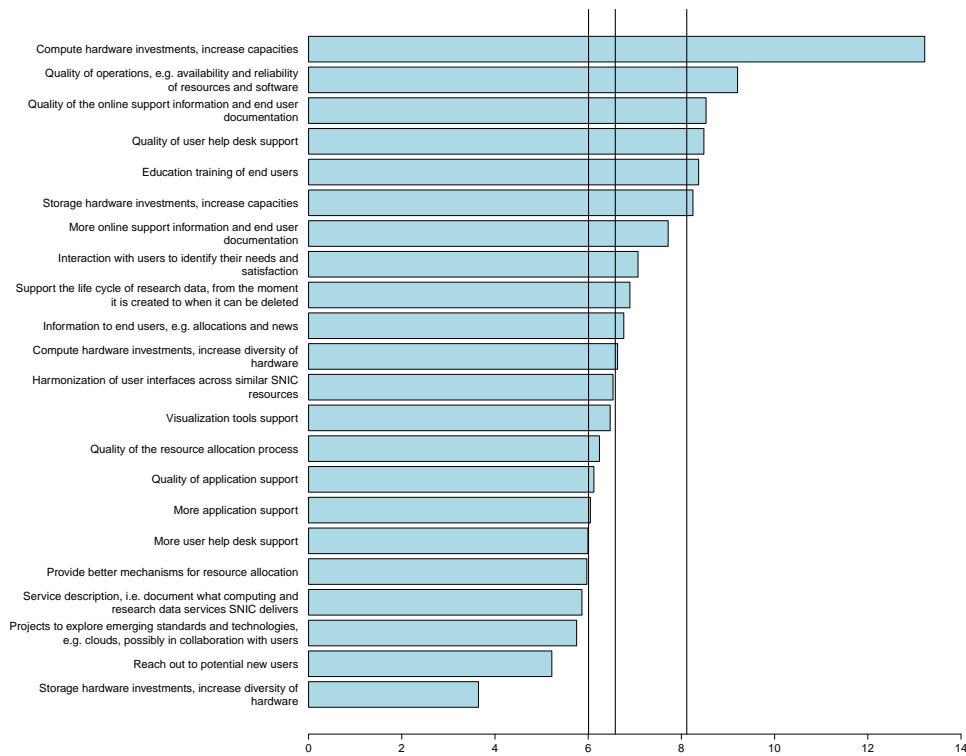

Figure 13: Respondent preferred focus areas for SNIC in the coming years. Out of the 159 respondents, 155 had marked their preferences in Q85 – Q106. Each user's preference ratings are normalised by their total number of rating points awarded. The vertical lines in the graph show the 25%, 50% and 75% quantiles from left to right. Note that the ordering of the questions in the graph does not follow that of the survey.

Q107: *In case there are specific reasons why SNIC should prioritize one or more of the above areas, please state this here. If you wish that SNIC prioritizes other areas that are not listed, please specify and motivate this also here.*

R17:
"SNIC should not try to be the general helpdesk for computers in Sweden, but focus on the parts that cannot be accomplished outside the national infrastructure.Just as for most other areas, I think SNIC would improve from focusing on their national task, and actually leave some local infrastructure challenges to other players."

R22:
"Simply more flexibility for storage and updated hardware, especially for computing nodes, will have the utmost impact in my research. "

R48:
"The remote visualization tool at LUNARC is useful as it allows me to do visualisation on larger data sets than I can on my desktop computer. The reason for this is that I can access much more memory on the remote service. I think it can be develop further and this is kind of remote visualization is what I refer to in "Visualization tools support"."

R52:
"More compute hardware capacity to UPPMAX, and possibly other small SNIC centres, making it possible for small projects to advance into medium size projects."

R53:
"Education/training of end users - would be very nice to have some very general scientific computing training (may be area-specific, like physics, chemistry, etc) but not software specific (not just linux, parallel programming, VASP, etc.)"

R57:
"Minimise down time generally, plus improve head node performance."

R68:
"Documentation is important for reproducibility! Reproducibility is important for science! We are scientists. :-)"

R82:
"The visualization tools I used like VESTA are working for me very slow. This cause me to avoide using it."

R91:
"To the last above: SNIC is our cloud. Do we need a cloud-cloud ? :-)"

R99:

"The main problem today is lack of hardware capacity. There is no need for more "diversity". The already existing kind of hardware is fine for 99% of the usage. In my field, we just need more core hours."

R120:

"Amount of research data in future will keep increasing, therefore more users will need advanced computing beyond their desktops, and SNIC will encounter new applications and use cases that will need new approaches and perhaps new technologies. Hence all the priorities."

R125:

"You should prioritise making the current file system stop hanging."

R127:

"My students are application experts so we don't use any of SNIC's applications support. "

R134:

"SNIC should definitely invest in more/larger CPU/GPU clusters. ERIK (LUNARC) has been my faithful workhorse ever since it started. The wall times are painfully short though. 48 hours is way too short. Luckily I can make use of checkpoints to restart simulations, otherwise I don't know what I would do. I do not know the exact reason for such short wall time, but I have a feeling it's due to the small amount of available nodes. More "thin" and "fat" nodes would be great, because it is clear that more and more users are starting to use the aforementioned cluster and it will get overcrowded soon enough."

R137:

"I am just working with uppnex."

R144:

"Lacking computing power is one of the main problems to climate science and therefore I think that the work with increasing that should be prioritized over other areas. "

R145:

"HPC must be #1 priority at all time"

R150:

"Our research is today limited by the allocation we have, and would be performed quicker and better with more allocation. We would prefer a decentralized commitment (e.g. on C3SE) rather than a centralization towards few larger centres because of the importance of personal service, user influence on hardware and software, and possibilities to own investments in nodes."

R155:

"To me the main points are: To maintain large resources available to users since this ultimately determines what projects are possible. To educate users. Summer schools in HPC programming is a great investment for many PhD students. "

R158:

"There is always a need for 'bigger and faster'. Extend allocation period to e.g. 3 years for Large applications. Provide "standard scripts" for codes that are provided in the centers (in cases where this is not already done), avilable as downloadable files. "
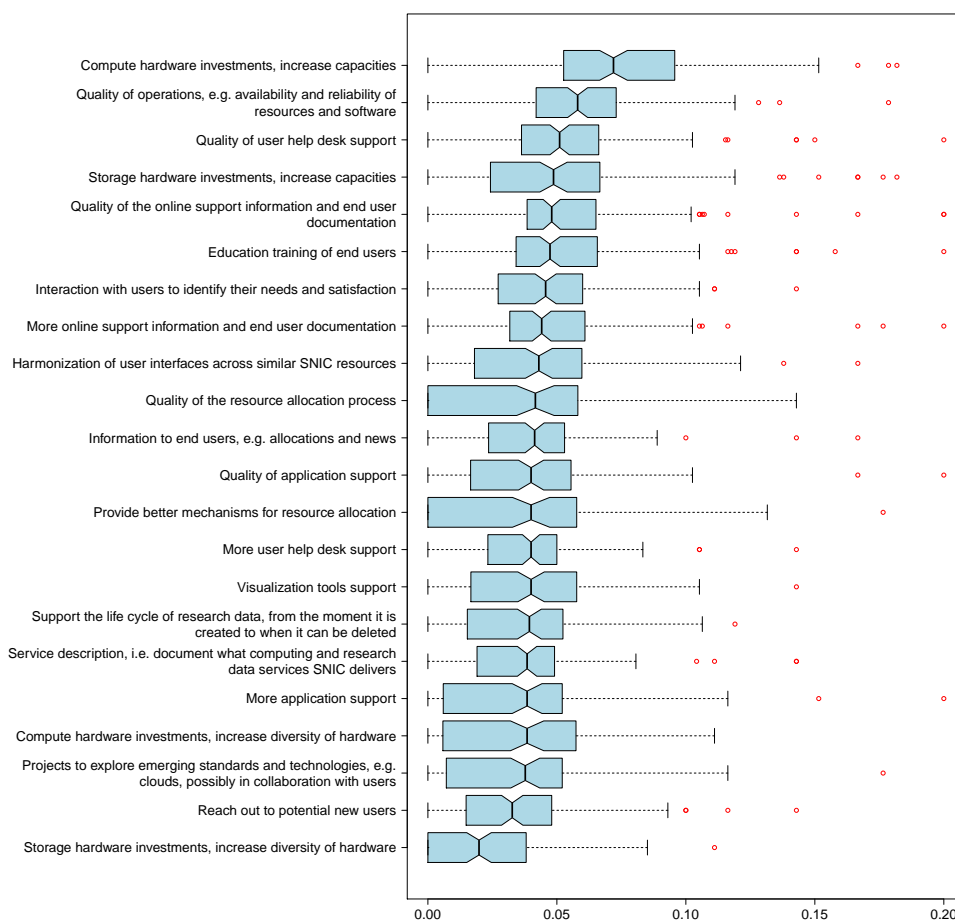

Figure 14: Notched box plot of the median sorted normalised average respondent scores for their preferred focus areas for SNIC in the coming years. The x axis data range is cropped from 1.0 to 0.2, and therefore some outliers in the data are not shown. Note that the ordering of the questions in the graph does not follow that of the survey.

8 Summary

This summary aims to describe the current situation within the SNIC community as well as the desired future direction for SNIC, as seen by the survey respondents. Since no deeper analysis of the data has been made, consider this summary an abridged version of the report with highlighted areas.

The summary to very little extent touch on the free form comments. This only reflects that they are difficult to summarise, not that they are uninteresting. Quite the contrary, they make a good read for the most part.

The "SNIC User Survey 2014 Report" abridged:

- All academic levels are represented by the survey respondents, who most commonly use the SNIC resources on close to a daily basis and have done so for quite a few years.
- All manner of job sizes between one and ~ 10k cores and job lengths are being run by respondents. There is a slight predominance for jobs sizes between a single node (all cores) job and up to 256 core jobs in terms of frequency. The same job size bracket is also predominantly used for long jobs, while short jobs are run mostly on fewer cores.
- The differences between centres is annoying to the respondents but not a show stopper for most. Higher quality documentation could probably make these differences less troublesome. Minimizing center differences with respect to job submission and software availability would make the highest impact here.
- Proximity to hardware is pretty much a non-issue, but proximity to the people running it is for many respondents.
- There seems to be a certain measure of unmet demand for large memory systems.
- For the future sizes and amount of SNIC compute resources it can be noted that almost as many respondents wanted larger and fewer systems as those who wanted the current mix to remain. A small minority answered that their interests were best met with more and smaller systems.
- Perhaps not surprisingly, there is no majority view among the respondents on how the best future development of SNIC compute resources would look like. Slightly more respondents would like to see a lower emphasis of Heterogenous computing than higher. Significantly more wants higher emphasis on Capability computing than lower. For Throughput computing it is basically a tie between preference for higher or lower emphasis. In all three categories, the most checked preference was that of keeping the same emphasis.
- The main data service survey respondents wanted SNIC to provide was long term research data storage.
- A high proportion of the respondents say they mainly store data that cannot

be regenerated, or is hard to regenerate. A high proportion of respondents also store their research data on center storage. Assuming a correlation between the two, it is natural that a highly ranked future focus for SNIC is to maintain a high quality of operations such as resource reliability and availability.

- The survey respondents are, by and large, very happy with the SNIC user support both in terms of quality and in form (e-mail). Nevertheless this is a top rated focus area for the future.

- The use, or role, of the Application experts is not clear to some respondents. However, those who have been helped by them are happy with the assistance they got. Some notable exceptions exist.

- Continuous effort work of application experts is rated almost twice as high as doing short and long term projects. By respondent preference, the application expert efforts should mainly consist of setting up and maintaining software development environments, deploying and testing widely used scientific software as well as do hands on code optimisation work and support individual researchers or research groups.

- The SNAC allocation process got a rating centered on "Satisfactory". A very common need expressed in the comments was that a mechanism to better deal with uneven workloads is highly desirable. Some interesting suggestions to deal with this were put forward. Some comments express a need to get more feedback when proposals are cut back or rejected.

- The top rated future focus areas for SNIC were:

  1. Compute hardware investments, increase capacities.

  2. Quality of operations, e.g. availability and reliability of resources and software.

  3. Quality of the online support information and end user documentation.

  4. Quality of user help desk support.

  5. Education, training of end users.

  6. Storage hardware investments, increase capacities.

  7. More online support information and end user documentation.