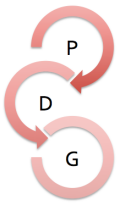


Report of 2015 UK National e-Infrastructure Survey of HEIs and Research Institutes

Martin Hamilton (Jisc/NeI PDG - editor)
Clare Jenner (UCL/DiRAC/NeI PDG)
Jacky Pallas (UCL/eMedLab/Farr/NeI PDG)
Alan Real (Leeds/N8/HPC-SIG)
Andrew Richards (Oxford)
Jeremy Yates (UCL/DiRAC/SKA/NeI PDG)

UK National e-Infrastructure Survey coordinated for Project Directors Group by Jisc

Contents

Executive Summary

This is the draft report of the third annual UK National e-Infrastructure (NeI) Survey. This is intended to inform the management and development of the UK's National e-Infrastructure for research and innovation. The NeI is comprised of publicly funded facilities and expertise that are made available to academic researchers and for industrial and third sector engagements. There is a full table of facilities available in Appendix J, including Archer, the UK's national service, and systems provided by the STFC Hartree Centre.

This survey report summarises the current state of the UK NeI, reporting on recent capital investments, and integration activities designed to make the NeI more effective and user centric. We make a number of recommendations based on the findings from this survey, which will help to place the NeI on a more sustainable footing and ensure that the National e-Infrastructure underpins UK research and innovation that is globally leading.

Recommendations:

1. Training materials and courses should be easily available. A service to host materials and advertise courses and to assign courses to particular levels of skill and knowledge would be very helpful in making sure users get appropriate training and are able to progress easily from level to level. HEIs and Research Domains should be encouraged to share what materials and resources they can with each other. Training has to be provided for trainers.

2. A Training Framework, possibly on an on-line Marketplace, needs to be created to allow our user base to find out and access the right training needed for their project.

3. RCUK should as a matter of urgency prepare a business plan for capital and operational resources to support HPC provision in the NeI, indicating to what extent the reduction in international competitiveness and loss of productivity during the period 2016-18 can be mitigated. This will be necessary in order to realise the ambitions of initiatives such as the Turing Institute.

4. The newly established RCUK Cloud Working Group should develop a strategy to establish the scope and potential of Cloud services for research and innovation. This should identify which parts of the NeI could usefully adopt hybrid Cloud technologies, and how Cloud technologies could assist the NeI in sharing resources between its different projects.

5. To leverage UK investments in EU and international e-Infrastructures such as the European Science Cloud, EGI and EUDAT, RCUK should report on the current level of UK participation in these projects and devise a strategy to ensure the UK both maximises use, and influences the design and management, of these infrastructures.

6. HPC-SIG has a key role to play in facilitating NeI providers' self-organisation into partnerships/consortia to build critical mass of knowledge, achieve operating efficiencies and maximise productivity, creating economies of scale without compromising diversity.

7. An economic sustainability model of the current NeI should be constructed as a matter of urgency and a sustainability strategy developed by the main participants of the NeI. Workshops will be held to develop this model and to develop partnerships in service provision. This is necessary to assist with planning and to allow comparison of service costs with Public Cloud providers.

8. E-Infrastructure providers need to be given confidence that there is a long term plan if they are to release funds that will underwrite staff roles beyond the lifetime of current projects, and to develop a career structure for staff working in "Research Operations" (ResOps).

9. RCUK should propose common research and innovation related metrics for all NeI providers to ensure a standard of reporting to allow the effectiveness, efficiency and diversity of NeI service provision to be assessed.

10. Service Providers should be encouraged to form partnerships and pool resources to deliver economies of scale in providing a full range of services across academic, industry and third sector organisations.

11. Increase training and update service management protocols appropriately, so that those using and managing NeI services are aware of the security requirements of their projects and their systems respectively. This should be done in a proportionate way that reflects data being processed and uses/users of the systems.

12. Explore the potential of using Jisc's VAT cost sharing group to reduce the friction of sharing facilities such as e-Infrastructure, and research equipment more generally.

13. The ELC should organise meetings between regional consortia/partnerships and National Projects with LEPs and initiatives such as the Northern Powerhouse and Midlands Engine to ensure that local e-Infrastructure needs are assessed.

14. IUK should work with regional consortia/partnerships and National Projects to develop the Catapults and work effectively with the research community and government. Examples include the Precision Medicine Catapult which will have its HQ in Cambridge and the Medical Technology Catapult which be located close to the Hartree Centre.

15. Industry should have improved visibility of NeI resources, e.g. via access to dashboards and portals developed by the NeI for its own use.

16. Regional and national facilities should have an industrial engagement strategy including metrics that permit the effectiveness of the strategy to be quantified.

17. Leverage existing and new RCUK-funded business development staff to support e-Infrastructure related elements of industry engagement, working with regional centres and local HEIs.

18. Encourage use of strategic funding such as Impact Acceleration Accounts to pump-prime business development activities within HEIs specifically for compute and data resources.

19. The new AAAI application SAFE+Assent and the secure access and secure data transport infrastructure Safe Share should be tested and productised in readiness to be rolled out to the NeI. A sustainability model is also needed for both these services.

20. End-to-end network connectivity diagnostics should begin to be undertaken between NeI providers and users' local research organisations to identify network bottlenecks and provide research organisations with a diagnosis report.

21. Common core applications for the UK Data e-Infrastructure in order to move and select data from distributed datasets should be developed and tested.

22. An RCUK metadata strategy to be developed to ensure that metadata is generated at the point of data creation, and that standard metadata query and analysis tools are provided to UK researchers.

Progress report on recommendations of 2014 NeI survey

1. The connections of NeI Service providers to the SJ6 backbone be evaluated and if necessary upgraded or separate links be provided.

2. Internal investment by institutions is required to ensure that internal campus networks remain fit for purpose. This needs to be communicated to HEIs via their Pro-VCs for Research.

Jisc have recruited a team led by Tim Chown to deliver an end-to-end network connectivity diagnosis service. From the Autumn of 2015 this team will work with NeI Projects and HEIs to test the efficacy of links from local institutions' services to the SJ6 network. These tests and a suggested investment strategy will be made to the participating institution. This should ease data congestion within the network infrastructure operated by the institution and increase research productivity.

3. A long term capital plan is required to ensure the future productivity of the NeI. This should be co-ordinated and carry on momentum established to create a holistic e-Infrastructure eco-system for the UK. Being able to plan will greatly increase the efficacy and efficiency of our systems and release resources for added value activities such as software engineering. Greater coordination has the potential to establish deeper and more valuable partnerships with the vendor community.

The UK NeI Community produced four key reports to address how we create an integrated holistic e-Infrastructure:

- The RCUK E-Infrastructure Roadmap, produced by the RCUK NeI Group, which put forward a coherent strategy to integrate and develop the UK National E-Infrastructure so that it can drive forward the continued development of a globally competitive research base within the UK.

  See **http://www.rcuk.ac.uk/RCUK-prod/assets/documents/documents/RoadmapforELC.pdf**

- The RCUK Data for Discovery Workshop Report, produced by the RCUK NeI Group, that produced recommendations for the formation of a UK National Data for Discovery Infrastructure.

  See **http://www.rcuk.ac.uk/RCUK-prod/assets/documents/documents/RCUK%20DataforDiscoveryWorkshopReport.pdf**

- Cloud Computing for Research and Innovation, produced by the NeI Project Directors Group, which produced a strategy for make use of cloud technologies and cloud providers to enhance e-infrastructure outcomes in

the UK.

See **https://www.scribd.com/doc/273829152/Cloud-Computing-for-Research-and-Innovation**

- Imagining the UK National Data e-Infrastructure, produced by the NeI Project Directors Group, which identified the core technologies required to power a future UK National Data for Discovery Infrastructure

See **https://www.scribd.com/doc/260531862/Imagining-the-UK-National-Data-Infrastructure-Recommendations**

The RCUK NeI Group (Chair: Morrell) continues to co-ordinate e-Infrastructure strategy amongst the Research Councils, Innovate UK, Jisc and the Met Office.

In addition the RCUK NeI Group has funded (£40k) the NeI Project Directors Group (Chair: Yates) so that it can better co-ordinate e-infrastructure integration activities. This has allowed the PDG to organise meetings and workshops on

- Authentication, Authorisation and Allocation Infrastructure (AAAI)
- Cloud Technology and Cloud Provision
- Core Technologies for a National Data Infrastructure
- Disruptive Hardware Technology and its effects on Software needs and development

RCUK and PDG members made inputs into 2016-2021 science and capital plan
https://www.gov.uk/government/publications/our-plan-for-growth-science-and-innovation

4. Jisc is well placed to continue coordinating efforts in the AAAI area, building upon existing work wherever possible.

Jisc announced the new Authentication Infrastructure, formerly known as Project Moonshot (see **https://www.jisc.ac.uk/assent**) to manage access to both web and non web services. This will form the basis of a single sign on service for the NeI.

5. DiRAC, GridPP and Jisc work together to produce a prototype that allows these NeI Projects to share resources. This tests the single-sign-on capability of the proposed common AAAI infrastructure.

Jisc, eMedLab, DiRAC, GridPP, The University of Oxford, The N8 Regional HPC Service and the Edinburgh Parallel Computing Centre are participating in the SAFE+Assent Project to produce a prototype AAAI service that can provide single sign on services for the NeI. This will include a project around secure access and federating with projects that are unable to use Assent such as GridPP and Elixir.

6. eMedLab, Farr Institute, ARDC, EBI/ELIXIR and Jisc work together to produce a secure AAAI infrastructure that protects the security of the sensitive people data. This tests the data security aspect of the proposed common AAAI infrastructure.

Jisc have funded a project Safe Share, £960k, which provides both a 2 factor Authentication infrastructure and secure data transport (see **https://community.jisc.ac.uk/groups/safe-share-project**). This does not include the Life Sciences project ELIXIR who have their own system, developed in Finland, but this will be able for federate with Safe Share/Assent.

The Safe Share project has an obvious role in providing secure access distributed data and data transfer available to other sectors in the economy such as Health and Advanced Manufacturing.

7. JASMIN, Sanger Institute, GridPP, SKA and DiRAC should plan to work together to explore the practicalities of this approach and show that resources within Research Domains can be configured into effective and efficient private clouds that allow researchers to run their workflows easily on a domain private cloud or on resources in another part of the NeI.

GridPP, STFC Scientific Computing Division, SKA and DiRAC have formed an umbrella project UK-T0 to develop this approach. This builds on the progress already made by projects such as the CERN CMS experiment, the MRC Medical Informatics CLIMB and eMedLab projects, the EBI's Embassy Cloud and JASMIN2.

SKA have begun a project with the UK software company Canonical (**http://www.canonical.com/**) to develop the OpenStack Cloud operating system for use in data intensive projects and applications; thus making OpenStack more attuned to the needs of UK data intensive projects.

8. Jisc to take this work (public cloud) forward, in collaboration with major e-Infrastructure service users and public cloud providers. However the costs and benefits of such access will have to be carefully measured and it is unlikely to be a solution for many of our problem sets in the next 2-3 years. Public Cloud Providers and NeI providers should exchange information concerning the problem types and sizes in the NeI, so that expectations are not unduly raised and that methodologies are built up that allow effective and economic use of Public Clouds.

Jisc have surveyed the CIOs of the UK's HEI Sector to assess Cloud usage in general in the HeI Sector and found that there were a range of issues preventing more widespread adoption of cloud including legal and regulatory aspects and lack of clarity over pricing and costing for "bog standard" enterprise IT applications. See:

https://www.jisc.ac.uk/news/uk-education-divided-in-its-adoption-of-the-cloud-14-jul-2015

Jisc continue to work with Public Cloud providers on access portals and egress onto the SJ6 backbone. Jisc are investigating the offering of a brokering service to allow more economic use of Public Cloud services.

The PDG produced a report Cloud Computing for Research and Innovation.

RCUK NeI Group have established a Cloud Working Group (Chair: Kershaw, RAL), to drive the development of private hybrid clouds in the NeI and improve access to private clouds

An OpenStack (the leading open source Cloud Technology) User Group has been formed, supported by the UK SME OCF, to accelerate the uptake of Cloud technologies by UK NeI providers.

9. The e-Infrastructure Leadership Council is encouraged to consider potential approaches such as greater regional collaboration supported by an element of matched funding for the HEIs' e-Infrastructure investments. This should be highlighted to Pro-VCs for Research.

We welcome the refreshed commitment for the E-Infrastructure Leadership Council under the new administration. Their role in providing strategic advice to the NeI community is valuable particularly as we grow our engagement with industry.

This survey report makes a number of recommendations on how the NeI can better support Industry and specialist areas such as Health.

10. There is now the opportunity to formally recognise and define the contribution of the National "Centres" to the Ne-I and to make sure they are adequately funded to carry out their individual missions.

This still needs to be done.

11. Funders should make applicants aware that it is permissible to apply for research software engineer time on grants, and that it is also appropriate to class these as research staff on grants where the work to be carried out involves a significant research and development aspect.

This has been done, although it will take time to work through the system as grant panels need to understand this work improves the productivity of research projects.

12. The work to raise the profile of the research software engineer should continue - in particular the role, value and potential career paths should be highlighted into submissions to the ELC and HEIs. The value of dedicated developer support at HEIs should also be highlighted to Pro-VCs for research and Directors of Research, as well as to successful funding models.

The EPSRC call for Research Software Engineering Fellowships was issued in May 2015, providing support of up to £3.7M for Research Software Engineers., and a network led by the Software Sustainability Institute which is open to all Research Software Engineers.

See **https://www.epsrc.ac.uk/funding/calls/rsefellowships/**

13. Training materials and courses should be easily available. A service to host materials and advertise courses and to assign courses to particular levels of skill and knowledge would be very helpful in making sure users get appropriate training and are able to progress easily from level to level. HEIs and Research Domains should be encouraged to share what materials and resources they can with each other. Training has to be provided for trainers.

14. A Training Framework, possibly on an on-line Marketplace, needs to be created to allow our user base to find out and access the right training needed for their project.

The Archer National Service has made places on its courses available to those not funded by EPSRC and NERC, provided researchers funded by those Research Councils also attend those courses. These courses are held now held at many sites around the UK.

The Archer National Service has produced its own Driving Test to check and improve the basic IT skills of its new users. Both DiRAC and Archer now demand such tests are taken.

The Software Sustainability Institute contains to focus on improving the skills of those entering the NeI via the HEI sector. These include train-the-trainer training sessions, software carpentry workshops and the provision of teaching materials.

DiRAC and Archer have produced a many core programming course and a programming course for the Intel Xeon Phi, which will be made available to the NeI. Both the SKA and SES5 have produced many core GPU programming courses.

DiRAC and Archer have produced a code optimisation course that will be made available to the rest of the NeI.

The Hartree Centre continues to offer its 3 week Summer School to the academic community and industry, as well as a full programme of training events aimed at industry.

Sanger and EBI offer training at their campus at Hinxton to over 1500 early career researchers each year.

The potential and practicality of an online training market place and training resources sharing infrastructure has still to be investigated. The BBSRC funded GOBLET (**http://mygoblet.org/**) and TeSS (**http://elixir-uk.org/training-platform**) projects should be investigated for potential transferability to other domains.

15. A process is put in place to make sure the on-ramp centres and the NeI work together so that SMEs can actually access NeI resources. This should include schemes to induce SMEs to use NeI infrastructure.

Innovate UK have set up the e-Infrastructure Special Interest Group to address this issue. Members of IUK and the KTN are members of the RCUK NeI Group and the PDG and meet regularly with other members of both Groups.

See **https://connect.innovateuk.org/vi/web/high-performance-computing**

E-Infrastructure capital investments and major initiatives

The 2014 Survey detailed ongoing investments in Big Data Analytics Projects in the areas of Health Informatics, Medical Informatics, Life Science Bioinformatics, Administrative Data Research, Analysis of Population and Business Data, Energy Efficient Computing, the Square Kilometre Array Project, Earth Observation, and Digital Arts and Humanities. This represented investment in excess of £300m in e-Infrastructure.

Many of these projects have (or about to) come on line in 2015 and contribute to the UK's research outputs.

Several new initiatives in data science and simulation have been announced. These strategic investments enhance the UK's ability analyse and model data in key economic areas such as Health and Society and Weather and Climate Change.

Project	Research Organisation	Amount / £M
Cognitive Computing and Data Science	The Hartree Centre STFC (with IBM Watson)	115 (BIS) 200 (IBM)
Alan Turing Institute	EPSRC, UCL, Edinburgh, Warwick, Oxford, Cambridge	42 (BIS) 25 (HEIs)
Genomics England	NHS and MRC	24
The Met Office	BIS	97
Big Data and the Information Economy	ESRC	75
TOTAL		**£378m** **£200m (IBM)**

There have also been several investments in local HEI provision, notably at UCL with £2m being an invested in a 185Gflop/s system. Southampton, Imperial and Oxford are also considering similar investments.

The Status of the Regional and National HPC Projects

Regional consortia

At a level above the local institute resources there exists a level of regional HPC resource. Five of these regional HPC centres were created with seed funding by EPSRC in 2011/2012. The current status of these regional centres varies and is summarised below.

Regional	Consortia HEIs	Status
SES5	Oxford, Cambridge, Imperial, Southampton, UCL	All services end July 2015. Loss of main UK GPU service. Imperial, UCL, Oxford and Southampton all have plans to replace/have replaced these services. Loss of collaboration around industrial engagement
N8	Leeds, Manchester, York, Sheffield, Liverpool, Durham, Newcastle, Lancaster	Main N8 Service hardware support ends in early 2016.
HPC Midlands	Loughborough, Leicester	Support until June 2016 for its Hera HPC service and they expect to continue to be operational until then.
ARCHIE-WeSt	Strathclyde, Glasgow, Glasgow Caledonian, West of Scotland and Stirling	Hardware support until 31/3/17. Hardware already 3.5 years old.
MidPlus	Warwick, Birmingham, Nottingham, Queen Mary University of London	MidPlus has support agreed until the end of March 2016

The National HPC Projects

Three projects received significant funding in 2011-2012 to set up baseline HPC resources in the UK to support simulation services for academia and industry.

- The **DiRAC HPC Facility**. This Facility became operational in Autumn 2012 and is now three years old. It supports research in theoretical particle physics, solar and planetary science, astrophysics and nuclear physics. It contains one of the UK's petaflop systems and clusters designed to carry out data intensive modelling and simulation tasks. This Facility will be uncompetitive from the Autumn of 2015.

- The **Hartree Centre**. This Facility became operational in Autumn 2012 and is now three years old. It supports innovative software development for Industry and provides simulation services for Industry. It contains one of the UK's petaflop systems and has systems designed to carry out data intensive computing. This Petaflop Facility will be uncompetitive from the Autumn of 2015.

- The **National Service Archer**. This Facility became operational in Autumn 2013 and is now two years old. It supports simulation for the engineering, physical sciences and natural environment research communities. It contains the largest the UK's petaflop systems and is to carry out compute intensive research. This facility has a planned five year lifespan, terminating in 2018.

- The three Facilities contain different hardware configurations that are tuned to solve a particular class of problems. Workloads from each Facility would not run optimally on another Facility and would interfere with the core mission of each Facility.

Bootstrapping the Alan Turing Institute

Whilst the Alan Turing Institute was still in the process of being set up as the NeI survey was being undertaken, we recognise that with no capital allocation for its own facilities the ATI will be wholly dependent on the NeI for its data science and compute requirements. Without a sustainable funding model for the NeI, projects like the Turing Institute will struggle to get off the ground, and may struggle to achieve their objectives – chief amongst which is to improve the productivity of the UK economy.

RCUK should as a matter of urgency prepare a business plan for capital and operational resources to support HPC provision in the NeI, indicating to what extent the reduction in international competitiveness and loss of productivity during the period 2016-18 can be mitigated. This will be necessary in order to realise the ambitions of initiatives such as the Turing Institute.

Who are the UK's national e-Infrastructure providers?

The HEIs remain the main provider of data and compute services to the National E-Infrastructure. 37 HEIs are providing services into the NeI. These range from the Edinburgh Parallel Computing Service, which provides Petaflop services to the National HPC service and DIRAC respectively and petascale research data services to the UK research community, to the University of Cranfield which provides services to engineering simulation projects.

These services support by the HEIs are the National Service Archer, the Research Data Facility, DiRAC, GridPP Tier 2 sites, the ESRC Administrative Data Centre and Administrative Data Centres, MRC and Charity funded Medical Informatics, the FARR Institutes Health Informatics, Life Science Bioinformatics, the SKA, the Regional HPC Centres, as well as local provision at HEIs. These cover the whole range of research funded by the Research Councils and HEFCE.

Jisc provides very high quality network services between research organisations via the Janet network.

The International, Charity and Research Council Central Laboratories provide specific large scale services in the areas of Genomics (Sanger Institute) and Bioinformatics (EMBL-EBI), support for central experimental facilities (STFC SCD provides services for Diamond, GridPP Tier 1, JASMIN2, ISIS, UK Observatories), Software Data and Innovation services for Industry (the STFC Hartree Centre).

The Met Office and the European Centre for Medium-Range Weather Forecasts (ECMWF) are the main provider of services to the weather forecasting and climate modelling communities. Both these Facilities provide petascale services to their user communities.

A new development has been the use of Private Cloud and Public Cloud services to provide resources to NeI users. This is wholly in the area of data for discovery area.

- Genomics England have rented resources from a UK based IT provider
- Several HEIs and the Medical Informatics project eMedLab have co-located in the Infinity Datacentre at Slough. Jisc assisted with this co-location project. An OpenStack private cloud is being put in place to manage these co-located services in an efficacious manner to increase research outcomes.
- An increasing number of HEI based projects are making use of Public Cloud provision and are making use of Amazon Web Services, Microsoft Azure and Google resources. The scale of this usage is very hard to determine; the main driver is probably mainly the use of virtualisation to allow workflows that were built on local workstations to run easily on large virtual clusters without the need for time consuming porting of workflows to clusters in the NeI.

- HEIs are making increase use of Cloud resources for back office and administrative services used by their organisations.

Recommendations:

The newly established RCUK Cloud Working Group should develop a strategy to establish the scope and potential of Cloud services for research and innovation. This should identify which parts of the NeI could usefully adopt hybrid Cloud technologies, and how Cloud technologies could assist the NeI in sharing resources between its different projects.

To leverage UK investments in EU and international e-Infrastructures such as the European Science Cloud, EGI and EUDAT, RCUK should report on the current level of UK participation in these projects and devise a strategy to ensure the UK both maximises use, and influences the design and management, of these infrastructures.

Sustainable funding for the NeI Ecosystem

The HEI Problem

The advantage of placing services in HEIs is that it places services in the same location as the research communities that use them, so allowing greater interaction between service provider and researchers.

It also provides an important local resource for researchers at HEIs and encourages the entry of new researchers and research areas into using large scale data for discovery and simulation services, by providing both local help and training.

HEIs are also all engaged in work with public sector, businesses and commerce. They are an obvious interface with SMEs.

The HEIs are the engine that will drive take up of these new services and therefore need to be suitably resourced. However this is difficult to achieve in a co-ordinated fashion. HEIs each have a particular research mission that they wish to fulfil and their 5 year plans will be markedly different from each other.

Now that the 2016-2021 science capital funding framework is in place and an E-Infrastructure Roadmap exists an opportunity presents itself to devise a sustainable funding framework for HEIs. This should recognise that:

1. fEC income into HEIs should allow HEIs to deliver cores services to their local researchers.
2. Capital monies from RCUK and other sources can be used to help to HEIs provide services for both national projects and specialist projects.
3. The existence of the NeI Survey, a funding framework and the roadmap should allow NeI providers to self-organise and
    - create partnerships/consortia to provide services
    - pool and aggregate resources via co-location or make use of cloud services
    - create services and the needed critical mass of expertise around particular service requirements or research areas
    - deliver a greater range of services, particularly around industrial and public sector engagement, research software engineering and training services.
4. Operational costs remain an issue in the current funding climate and HEIs can assist with meeting this issue if capital can be used to fund their infrastructure.

Recommendation:

HPC-SIG has a key role to play in facilitating NeI providers' self-organisation into partnerships/consortia to build critical mass of knowledge, achieve operating efficiencies and maximise productivity, creating economies of scale without compromising diversity.

Non HEI Providers

The Research Councils and other research organisations have made/will be making their capital plans and planning for the associated operational costs.

There is merit in having closer co-operation and knowledge sharing between those engaged in providing services for our largest projects. This can lead to reductions in project operational costs and make projects more productive.

Recommendations:

An economic sustainability model of the current NeI should be constructed as a matter of urgency and a sustainability strategy developed by the main participants of the NeI. Workshops will be held to develop this model and to develop partnerships in service provision. This is necessary to assist with planning and to allow comparison of service costs with Public Cloud providers.

E-Infrastructure providers need to be given confidence that there is a long term plan if they are to release funds that will underwrite staff roles beyond the lifetime of current projects, and to develop a career structure for staff working in "Research Operations" (ResOps).

RCUK should propose common research and innovation related metrics for all NeI providers to ensure a standard of reporting to allow the effectiveness, efficiency and diversity of NeI service provision to be assessed.

E-Infrastructure as a shared service

The publication of Sir Ian Diamond's report "Efficiency, effectiveness and value for money" has served to emphasise the impressive gains that the university sector has made in improving equipment sharing and engagement with industry. The Diamond report included examples of e-Infrastructure equipment sharing (SES, UK Data Service) and highlighted the value of a cross-research council interest group to promote asset sharing.

Further big data investments by the Research Councils have centred on shared resources in regional foci such as N8 (Leeds Advanced Research Computing) and Wales (Farr/ADRC-W). MRC investments in medical bioinformatics have resulted in development of the shared eMedLab cluster (UCL, QMUL, LSHTM, Crick, Sanger, and EBI). This equipment is located in the Jisc Shared Datacentre together with MRC- and NIHR-funded equipment belonging to Imperial College and KCL. Relocation of compute and storage from inadequate premises in London to a modern energy-efficient datacentre allows HEIs to free up space in central campuses and save money in power costs.

This concentration of co-located resources is largely a result both of large capital investments with little matching operational costs, but also because there is a small pool of people with the technical skills and knowledge required to operate these large compute and data storage systems*. Efficiency through co-location is not just about saving money but also a consequence of the lack of suitable skilled staff to build and support systems.* Co-location and partnership is needed to provide effective service provision to academic and industrial users.

This survey has highlighted the large use of services by the NHS and NHS related projects. This is chiefly a consequence of new investment by the MRC and DoH leading to these new communities interacting with the NeI. These would not have responded to the 2014 survey.

The NeI is already working to provide secure access and secure data services e.g. Safe Share for these new communities.

Jisc has established what is believed to be the largest VAT cost sharing group in the UK, with over 250 institutions participating: **https://www.jisc.ac.uk/about/vat-cost-sharing-group**.

Recommendations:

Service Providers should be encouraged to form partnerships and pool resources to deliver economies of scale in providing a full range of services across academic, industry and third sector organisations.

Increase training and update service management protocols appropriately, so that those using and managing NeI services are aware of the security requirements of their projects and their systems respectively. This should be done in a proportionate way that reflects data being processed and uses/users of the systems.

Explore the potential of using Jisc's VAT cost sharing group to reduce the friction of sharing facilities such as e-Infrastructure, and research equipment more generally.

Industry

Engagement with industry is still largely focused mostly in a few sectors - advanced materials and manufacturing, energy and environment, and life sciences. Engagement is driven by the type of industry located close to the NeI facility. Local Enterprise Partnerships, with their specific expertise in certain industry sectors, should provide an ideal route to build relationships with local businesses specifically SMEs.

In particular and element of NeI infrastructure should be part of the infrastructure planning by large scale regional initiatives such as the Northern Powerhouse, Midlands Engine, Wales and Scotland. Examples of this would be:

- A northern Jisc Shared Data Centre
- Improved connectivity for business and academia outside of the SE.

Recommendations

The ELC should organise meetings between regional consortia/partnerships and National Projects with LEPs and initiatives such as the Northern Powerhouse and Midlands Engine to ensure that local e-Infrastructure needs are assessed.

IUK should work with regional consortia/partnerships and National Projects to develop the Catapults and work effectively with the research community and government. Examples include the Precision Medicine Catapult which will have its HQ in Cambridge and the Medical Technology Catapult which be located close to the Hartree Centre.

Key findings from the Dowling report (BIS 2015) are mirrored in the responses to the survey. Effective brokerage between industry, particularly SMEs, and service providers is critical. HEIs in particular lack resource to identify commercial opportunities in contrast to the large/specialist and regional facilities (the Hartree Centre has 5 dedicated staff for business development). Where this has worked effectively, it has been due to partnership with institutional business development teams. One example is an analysis of the opportunities and barriers to SME use of regional compute resources, published by the EPSRC-funded Centre for Innovation in partnership with UCL Advances. A further challenge is the lack of awareness on the part of industry of the availability and type of resources as part of the NeI.

The Dowling report also noted that VAT was a particular obstacle to collaboration and innovation between academia and industry and recommended that “the government needs to address the issue of VAT on shared facilities as a matter of urgency”.

Recommendations:

Industry should have improved visibility of NeI resources, e.g. via access to dashboards and portals developed by the NeI for its own use.

Regional and national facilities should have an industrial engagement strategy including metrics that permit the effectiveness of the strategy to be quantified.

Leverage existing and new RCUK-funded business development staff to support e-Infrastructure related elements of industry engagement, working with regional centres and local HEIs.

Encourage use of strategic funding such as Impact Acceleration Accounts to pump-prime business development activities within HEIs specifically for compute and data resources.

Examples of industry collaborations

The EPSRC HPC Midlands supercomputing centre of excellence is providing access to £60M of supercomputing equipment to Rolls Royce, facilitated by a brokerage scheme developed with Jisc.

The Sanger Institute also has collaborations, and provides scientific IT for, on-site spin-out companies such as 14m Genomics and Congenica. The Genome Campus (Hinxton) is building a "Biodata Innovation Centre" next year, massively expanding, with space for 250 people working on industrial/commercialisation projects. The Sanger Institute will be providing the IT infrastructure for the Centre.

The Square Kilometre Array Science Data Processor project houses the SKA Open Architecture Laboratory in which IT hardware and software companies can work directly with the SDP team to develop technologies for the SKA Science Data Processor. Highlights include low energy compute clusters for data processing (ARM), very fast and cheap disk arrays for data intensive computing (DELL) and using the Cloud operating system Openstack to manage the SDP workflows (Canonical).

OCF, a UK SME, has supported an OpenStack user group to accelerate uptake of cloud technologies by the NeI community, including eMedLab, MRC CLIMB and the Crick Institute.

The UK government has committed £113M to the future of the Hartree Centre. This in turn has leveraged up to £200M of technology and onsite expertise from IBM. Further case studies are available from the Innovate UK e-Infrastructure Special Interest group.

The UK now hosts 7 Intel Parallel Computing Centres, which are designed to drive the development of applications and libraries on many core systems. This is the largest number of Centres outside of the USA and

DiRAC has the largest number of Centres of any project in the world. Activities vary from the porting of applications, building new applications, the development of training resources, data imaging and data analysis, heterogeneous architectures, new maths libraries for many core processors and the fine gained management of parallel jobs. These activities will allow the UK to be among the first to take advantage of many core technologies as it matures.

See **http://www.intel.co.uk/content/www/uk/en/processors/xeon-phi/intel-parallel-computing-centers-overview-video.html**.

DiRAC hosts three of these Centres at the Universities of Edinburgh, Durham and Cambridge (DAMTP). Centres are also located at the Hartree Centre, the University of Bristol, the Edinburgh Parallel Computing Centre and most recently Imperial College.

To give a flavour of the work done a summary of the projects carried out by DiRAC is given below. These show that by working on beyond the edge technologies with firms like Intel it is possible to produce these kinds of innovative low level tasks/libraries. These projects also support a mix of traditional HPC and new data science applications.

1. Developed the GRID data parallel library, and code stencils, to allow easier use of matrix methods in parallel code on many core systems (Peter Boyle, Edinburgh)
2. Tested and developed many core visualisation software using Xeon Phi many core systems - OSPRay (Paul Shellard, Cambridge))
3. Developed data analytics and data modelling workflows using a heterogeneous processor (CPU) and many core Xeon Phi system which was built in the UK (Paul Shellard, Cambridge))
4. Developed the fine grained parallel task management QUICKSCHED library, which makes poorly scaling parallel code scale much better. It manages the so-called load balancing problem and is really needed if we want codes to function at petascale (Richard Bower, Durham).
5. Applying novel statistical acceleration techniques to Bayesian Methods and Markov Chain Monte Carlo – thus greatly reducing the time searching parameter space (Mark Wilkinson, Leicester) and improving quantifiable uncertainties in important classes of simulations.

Appendix A - List of surveyors and acknowledgements

Susan Morrell (EPSRC)
Martin Hamilton (Jisc/NeI PDG)
Clare Jenner (UCL/DiRAC/NeI PDG)
Jacky Pallas (UCL/eMedLab/Farr/NeI PDG)
Alan Real (Leeds/N8/HPC-SIG)
Andrew Richards (Oxford)
Jeremy Yates (UCL/DiRAC/SKA/NeI PDG)

Appendix B - Who will receive this survey

The National e-Infrastructure Project Directors Group
RCUK National E-Infrastructure Group
BIS E-Infrastructure Leadership Council.

Appendix C - List of respondents

ARCHIE-WeSt
Cardiff University
Cloud Infrastructure for Microbial Bioinformatics – MRC CLIMB - Birmingham, Cardiff, Swansea, Cardiff
Cranfield University
DiRAC @ Durham University
DiRAC @ EPCC
DiRAC @ University of Cambridge
DiRAC @ University of Cambridge (DAMTP)
DiRAC @ University of Cambridge (HPCS)
DiRAC @ University of Leicester
Durham University

EMBL-EBI
eMedLab (MRC)- UCL, QMUL, LSHTM, Crick, Sanger, EBI
EPCC
Farr North, Health eResearch Centre
HPC Midlands
HPC Wales
Imperial College London
King's College London
Lancaster University
Loughborough University
N8 HPC
NBI Partnership Ltd
Norwich Bioscience Institutes (TGAC, JIC, IFR, TSL)
Queens University of Belfast
SES/CFI
STFC Hartree Centre
STFC Scientific Computing Division
The Francis Crick Institute
The Institute of Cancer Research
The University of Birmingham
The University of Nottingham
The University of Sheffield
UVRI/MRC Medical Informatics Centre
University College London
University of Aberdeen
University of Bath
University of Bristol
University of Cambridge
University of Edinburgh
University of Exeter
University of Glasgow
University of Leeds
University of Leicester
University of Liverpool
University of Manchester
University of Oxford
University of Portsmouth
University of Southampton
University of St Andrews
University of Sussex
University of Warwick
Wellcome Trust Sanger Institute

Appendix D – Service management: The survey questions

Q1 Organisation name
Q2 Organisational unit
Q3 Email address
Q4 Job title

Budget

Q5 Is your HPC budget ring fenced, or do you have to make a fresh case for support each time?
Q6 Is your primary budget for HPC CAPEX or OPEX?
Q7 Are you directly charged for power and cooling or other Estates costs?
Q8 Do you capitalise Estates costs, or pay them from your OPEX budget?
Q9 Tell us about your data centre(s) (select all that apply)
Q10 Data centre PUE (if known)
Q11 Data centre power draw, if known

Project management

Q12 How is the project management needed to implement and deliver new projects provided at your institution?
Q13 How are Project Management (New Service Implementation and Delivery) resources calculated?
Q14 Do you work out the resources needed in terms of FTE?
Q15 Which of the following best apply to your HPC/Big Data/Research Computing/Scientific Computing Activity?
Q16 Do you use metrics to work out the Project Management resources needed? [it is assumed these are charged against the Capital Cost of the Project]
Q17 Are they... [tick boxes for percentage range of capital cost due to Project Management]
Q18 Which of the statements best describe how you arrive at the Project Management resources needed to deliver and implement the project?

Staffing

Q19 How many FTE support your HPC activity?
Q20 FTE count if >4
Q21 How many women are involved in HPC service provision at your organisation?
Q22 What model of HPC provision has your organisation adopted?
Q23 Breakdown of effort by FTE
Q24 Women in HPC service management (FTE)
Q25 What is your approach to training? Please tick all that apply
Q26 URL for further information on training if applicable
Q27 How would you prefer to pool training across the sector?

Q28 What proportion of your researchers are in self-supporting research groups?
Q29 What are the major training challenges and which do you address locally?
Q30 Other training challenges
Q31 Other forms of support. Does your organisation do any of the following?
Q32 Are you able to provide HPC case studies for (e.g.) HPC-SIG website, RCUK and Innovate UK e-Infrastructure SIG?

Cloud and Shared Services

Q33 Which statements best characterise your organisation's use of "cloud" services?
Q34 Use of cloud services - further information
Q35 Which statement best characterises your organisation's approach to HPC as a shared service?
Q36 Which areas of HPC provision might you share?

Research data management

Q37 Do you have a Research Data Management policy?
Q38 Does your organisation have a Research Data Management service?
Q39 URL for further information about Research Data Management at your organisation if applicable
Q40 What technologies have you deployed that can be regarded as being for Data Exploration ("Big Data") activities?

Academic impact

Q41 What are your Key Performance Indicators to measure academic impact?
Q42 Refereed and conference papers published
Q43 URL listing these publications if applicable
Q44 Does your organisation produce an Annual Report?
Q45 Does your organisation publish Research Highlights?

Who uses your service?

Q46 Do you provide HPC services beyond your immediate organisation?
Q47 What services do you provide to third parties?
Q48 How many HEIs research groups use your services?
Q49 Public and third sector organizations
Q50 What percentage of your users are from industry?
Q51 Of industry users, what proportion are SMEs?
Q52 What percentage of system time is being used by SMEs?
Q53 Are you taking any specific steps to increase SME uptake?
Q54 What sectors are being served by your industrial users?
Q55 Please use this space for any further information you would like to provide

Appendix E – Service management: Summary of the survey data

Where feasible we have tried to break down survey responses by large and specialist facilities, regional centres, and higher education institutions. Even then, many HEIs clearly operate on a shoestring e.g. with minimal staffing, whereas others have large teams working on scientific computing - so it is not trivial to draw comparisons between the three groups. We have included HPC Wales as a "regional" facility because of its primary focus on Welsh researchers and industry. It should be noted that HPC Wales was not part of the EPSRC regional supercomputer centre initiative.

Budgeting model - The vast majority of respondents (73% of those who answered this question) indicated that their primary budget for e-Infrastructure was capital (CAPEX). However, whilst 63% stated that they were not charged for power and cooling or other Estates costs, 28% of respondents indicated that they paid Estates costs from OPEX - implying that some elements of the service provision beyond salaries were being covered via OPEX.

Third party data centres - The only third party data centre that respondents reported using was the Jisc Shared Data Centre, in use by 8% of institutions responding. Of those running in-house data centres, most stated a PUE of below 1.5. Several respondents indicated that they were only able to report data centre power draw for a building or an entire data centre which their facility was part of.

Project management approach - It was rare for respondents to call on third party project management services for new service implementation and delivery, with this almost always being done in-house, sometimes (13% of cases) via dedicated project managers based in IT Services departments. Respondents had a range of approaches to calculating their resource requirements for a major e-Infrastructure project, ranging from highly formalised to "back of the envelope".

68% of respondents stated that they calculated staffing costs for their implementation projects. It was not unusual (36% of respondents) for institutions to bid for additional staffing as part of an e-Infrastructure project proposal. Only one institution used metrics to work out project management resources needed, and 46% of respondents agreed with the statement that "we don't calculate these resources, it just happens".

Staffing - 24% of respondents (including HEIs and large and specialist facilities) stated that they had two or less full time equivalent employees supporting their e-Infrastructure activities. Conversely, 9 out 25 HEIs responding (36%) had teams of four or more employees working in scientific computing. One HEI reported 23 FTEs working on its services, but these larger staffing complements were typically associated with regional and large/specialist facilities.

Women in HPC - the survey showed that there were 55 women involved in service provision across all of the facilities, mainly working part time. The situation was particularly disappointing in HEIs, with only 14 women involved in service provision in institutions. Many respondents stated that they were unsure how many women were involved in their services, which is odd given the small team sizes at a large proportion of HEIs. A separate survey has been conducted of ARCHER and DiRAC service users, to gauge the situation from the user's perspective.

Training - 36% of respondents reported that they carried out introductory training courses. This covers just 40% of HEI respondents. Only 24% of HEIs responding operated an advanced training course, and only 34% of HEIs actively promoted training available from other sources e.g. ARCHER and application vendors. Only two institutions provided training as part of a Doctoral Training Programme, and 3 as part of a Centre for Doctoral Training. However, there was widespread interest in sharing training materials freely online (73% of respondents) and/or creating a marketplace for training materials and delivery providers (34% of respondents).

Skills - 10 out of 25 HEIs responding (40%) reported that less than half of their users were in self-supporting research groups. HEIs, regional centres and large/specialist providers all saw a range of training challenges rather than a single key issue - a change from the 2014 survey where Linux orientation emerged as a huge problem. We will follow this development up with HPC-SIG members to see whether the requirement for Linux skills has diminished, or it is becoming more common for new service users to already be familiar with Linux.

Community building - Institutions reported that they were augmenting training with a range of additional community and expertise building approaches, including informal networking for service users (68% of HEIs), wikis or on line forums (48% of HEIs), and public speaking at researcher focussed events (40% of HEIs) or dedicated scientific computing events (36% of HEIs). Many service providers (24% of HEIs and 50% of large and specialist facilities) were also able to provide case studies of how their service had been used.

Research Data Management - 76% of respondents said that they had a Research Data Management policy, with 4 HEIs replying that they were either unsure about this or did not have a policy at present. 51% of respondents had a pilot or production RDM service at the time of responding to the survey, with 5 unsure.

Academic impact - common Key Performance Indicators were the number of users of the service (63% of respondents), the number of research papers published (61%), the number of projects run on systems (51%) and the amount of research grant funding brought in (51%). However, only a handful of respondents (e.g. 4 HEIs) were able to list the numbers of published papers that had used their system, and only a small number of respondents were producing lists of published papers, an annual report or research highlights.

Cross sector and industrial impact - many respondents were involved in inter-institutional collaborations such as the EPSRC regional centres, the Farr Institute, GridPP and the Medical Bioinformatics Initiative. 51% of respondents, including 32% of HEIs, provided compute and storage services to third parties. 27% of respondents worked with the National Health Service. Industrial use of national e-Infrastructure facilities was largely confined to a small number of providers, with some HEIs outstripping large and specialist and regional centres. Just 7 providers reported that they were working with Small to Medium size Enterprises (SMEs).

Appendix F – Service management: Full break down of survey data

Budget

Q5 Is your HPC budget ring fenced, or do you have to make a fresh case for support each time?

	HEI	Large / Specialist	Regional	Grand Total
Fresh case to Organisation	12			12
Partly ring fenced	8	2	1	11
Fresh case to External Funders		5	1	6
Wholly ring fenced	4	2		6
Other	1	3	1	5
Grand Total	25	12	3	40

Q6 Is your primary budget for HPC CAPEX or OPEX?

	HEI	Large / Specialist	Regional	Grand Total
CAPEX	18	9	2	29
Other	3	2	1	6
OPEX	4	1		5
Grand Total	25	12	3	40

Q7 Are you directly charged for power and cooling or other Estates costs?

	HEI	Large / Specialist	Regional	Grand Total
No	19	4	2	25
Yes	5	5	1	11
Other		3		3
Don't know	1			1
Grand Total	25	12	3	40

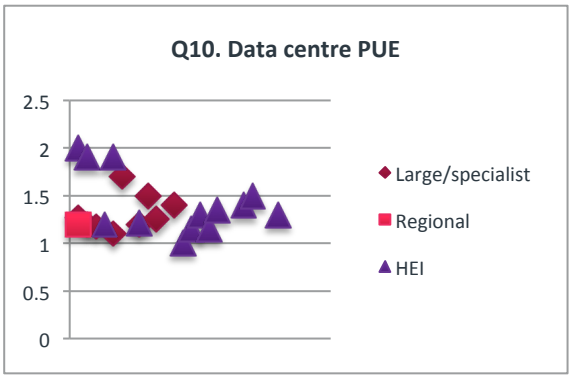
Q8 Do you capitalise Estates costs, or pay them from your OPEX budget?

	HEI	Large / Specialist	Regional	Grand Total
	19	7	2	28
OPEX	6	4	1	11
Other		1		1
Grand Total	25	12	3	40

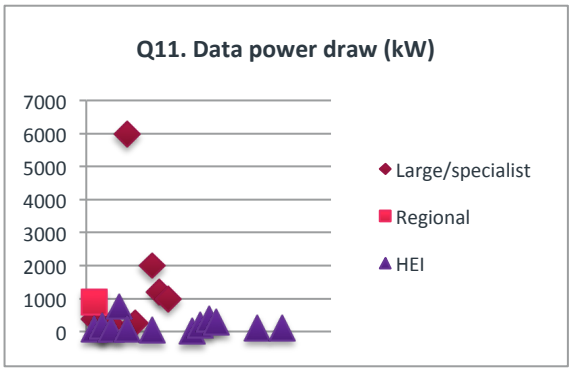
Q9 Tell us about your data centre(s) (select all that apply)

	HEI	Large / Specialist	Regional	Grand Total
We operate in-house data centres	24	12	3	39
We use the Jisc Data Centre	1	2		3
We use a third party data centre				
Grand Total	25	14	3	42

Q10 Data centre PUE (if known)



Q11 Data centre power draw, if known



Project management

Q12 How is the project management needed to implement and deliver new projects provided at your institution?

	HEI	Large / Specialist	Regional	Grand Total
We do some aspects of the project management using our in-house systems teams and the rest is provided by the successful vendor	12	4	3	19
We do ALL aspects of the project management using only our in-house systems teams	6	7		13
Our central IT dept. has project management staff and we can use them if we need to	4	1		5
We pay external contractors to handle ALL aspects of the project management	1			1
Grand Total	23	12	3	38

Q13 How are Project Management (New Service Implementation and Delivery) resources calculated?

- Supported by an STFC grant and are bundled with the overall DiRAC project
- Project Management resources are calculated on a needs basis, and currently we require one FTE, which is funded from two different sources
- We make an estimate of all staff effort (FTEs) for each project, at the project proposal stage.
- We use our 25 years experience when writing proposals and tenders to provide appropriate resources for each project

and service.

- We are at the initial start-up phase, so 100% of the time has thus far been on new service implementation and delivery.
- By the Business & Delivery teams, based upon an extensive spreadsheet
- Based on our experience and expertise and the customer requirements
- Based on standard FTE costs
- Very informally; for each procurement, one of the HPC architects will take on project management responsibility for the new service.
- Each Project and associated services are subject to an initial review by the Operational and User Services Group
- Project management costs for the implementation of a new HPC service are included in the system procurement
- No direct charging since these are internal resources
- It is based on the size of the contract and estimated deployment time
- Day rate based on actual costs blended across a mix of core permanent staff and contractors
- Formal Project Management is only required on an ad hoc basis so the resource calculations are similarly done on an ad hoc and case by case
- As part of our project delivery processes.
- Advise from in-house project delivery teams, typically 1 PM per major project.
- Part of the Sys Admins job
- Resources calculated internally

Q14 Do you work out the resources needed in terms of FTE? [It is assumed these are charged against the Capital Cost of the Project]

	HEI	Large / Specialist	Regional	Grand Total
No	6	1		7
Yes	6	8	1	15
Grand Total	12	9	1	22

Q15 Which of the following best apply to your HPC/Big Data/Research Computing/Scientific Computing Activity?

	HEI	Large / Specialist	Regional	Grand Total
They are calculated in terms of required FTE and some of this is tensioned against our overall staff resource allocation and the rest is requested in the Project Tender.	2	5	1	8

They are calculated in terms of required FTE and this is tensioned against our overall staff resource allocation.	3	3		6
Unsure	2	2		4
They are calculated in terms of required FTE and some of this is tensioned against our overall staff resource allocation and some is requested from our central IT dept.	1		1	2
They are calculated in terms of required FTE and all of this HAS to be provided by our central IT dept.	1			1
They are calculated in terms of required FTE and all of this is requested from our central IT dept.	1			1
Grand Total	10	10	2	22

Q16 Do you use metrics to work out the Project Management resources needed? [it is assumed these are charged against the Capital Cost of the Project]

	HEI	Large / Specialist	Regional	Grand Total
No	9	2	1	12
Yes	1			1
Grand Total	10	2	1	13

Q17 Are they...

Less than 5% of the total capital costs
Between 5 and 9.99% of the total capital costs
Between 10 and 14.99% of the total capital costs
Between 15 and 19.99% of the total capital costs
Over 19.99% of the total capital costs
Unsure

One institution responded to say that their project management costs were less than 5% of total capital costs.

Q18 Which of the statements best describe how you arrive at the Project Management resources needed to deliver and implement the project?

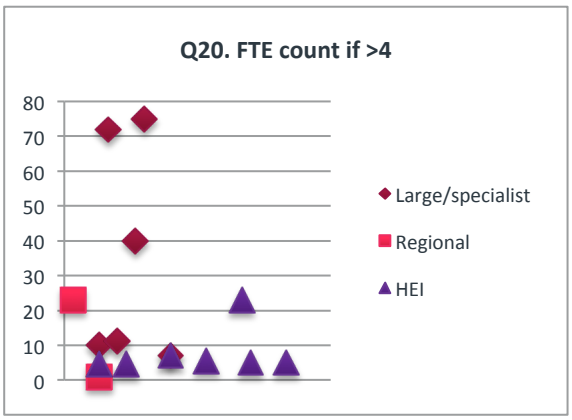
	HEI	Large / Specialist	Regional	Grand Total
We don't calculate these resources - it just happens	4	1	1	6
We don't calculate these resources because these resources are already included in our overall service delivery plans.	4	1		5
Already included	1			1
Unsure	1			1
Grand Total	10	2	1	13

Staffing

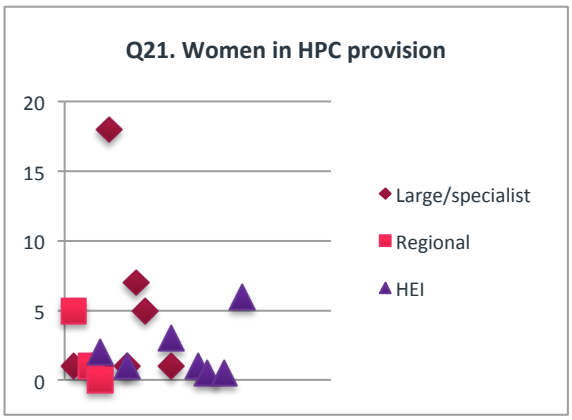
Q19 How many FTE support your HPC activity? Please choose the nearest answer and enter the exact figure below if >4

	HEI	Large / Specialist	Regional	Grand Total
1	2	1		3
1.5	3	1		4
2	2			2
2.5	5	1	1	7
3	2	3		5
3.5	1			1
4	3		1	4
>4	6	5		11
Grand Total	24	11	2	37

Q20 FTE count if >4



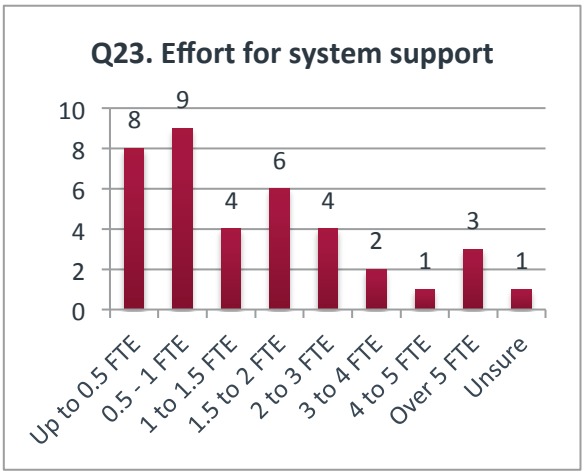
Q21 How many women are involved in HPC service provision at your organisation?

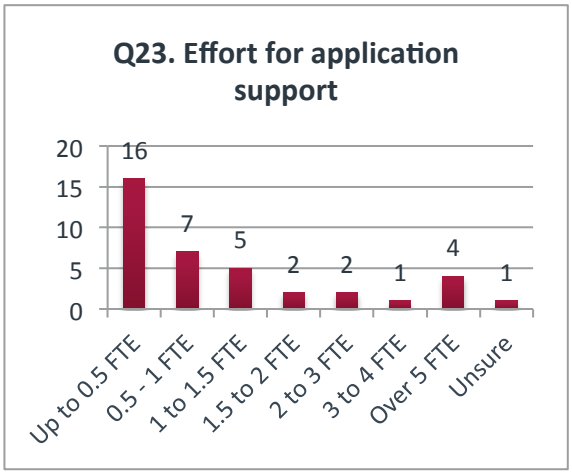


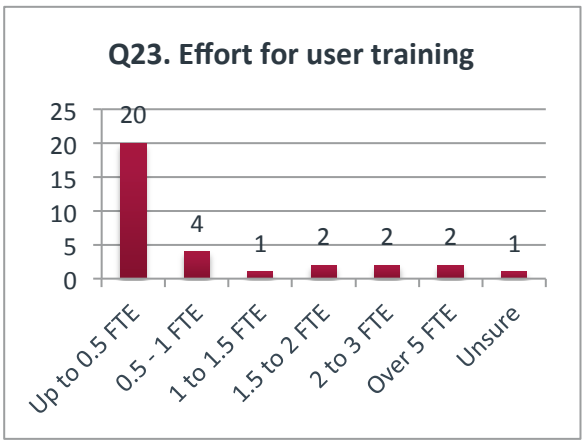
Q22 What model of HPC provision has your organisation adopted?

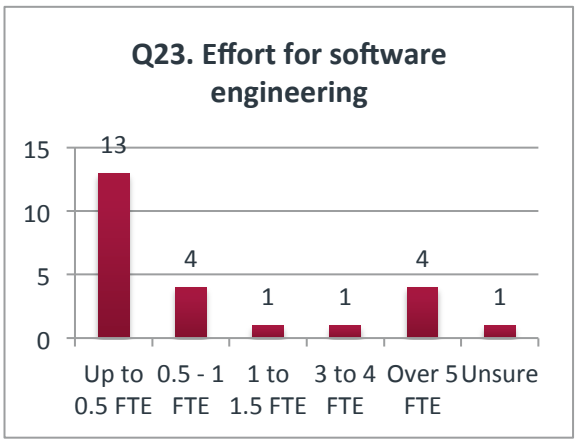
	HEI	Large / Specialist	Regional	Grand Total
Central IT department	2	3		5
Devolved to Faculties and Schools			1	1
Government research establishment		2		2
Independent centre		2		2
Mixture of devolved and central	4			4
Other	1			1
Regional or national facility		1	1	2
Grand Total	7	8	2	17

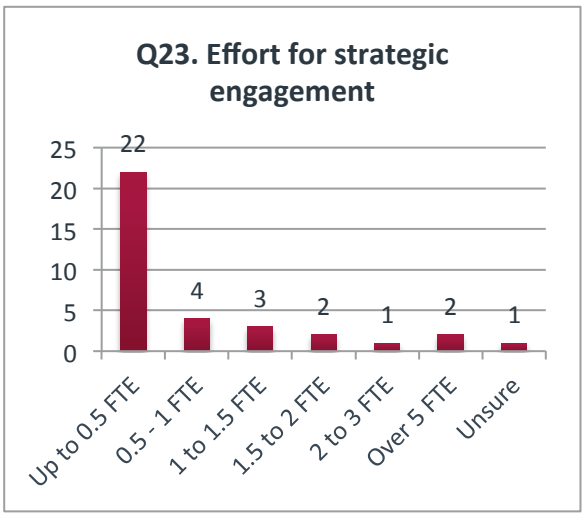
Q23 Breakdown of effort by FTE

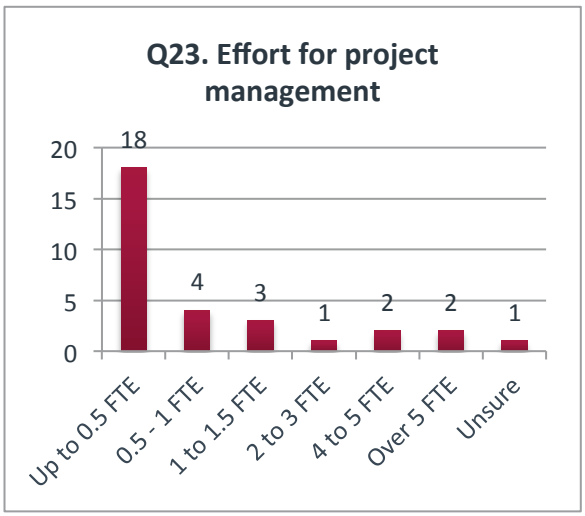


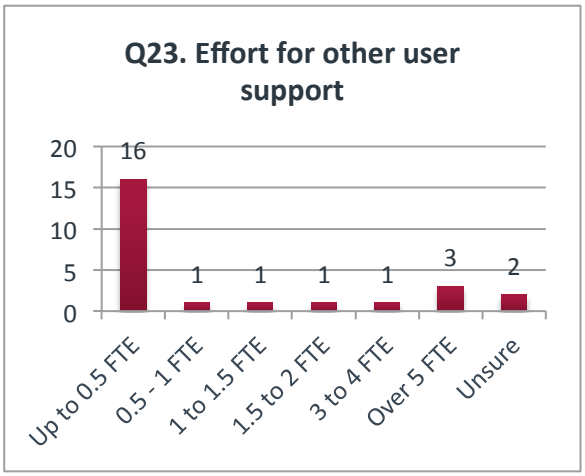




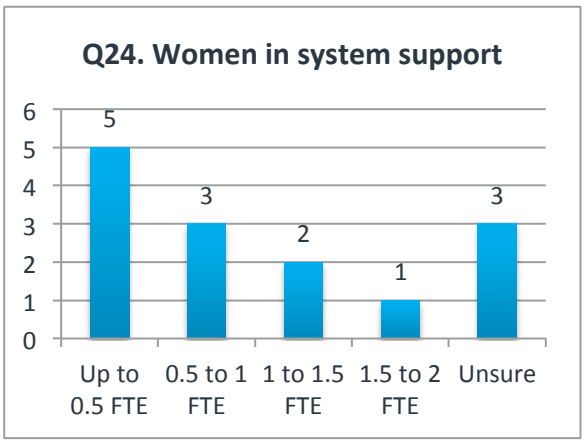


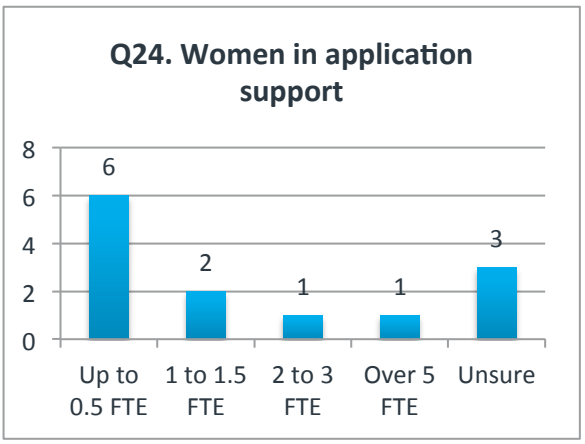


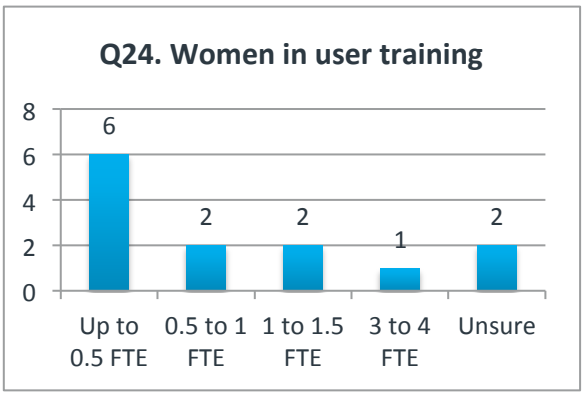


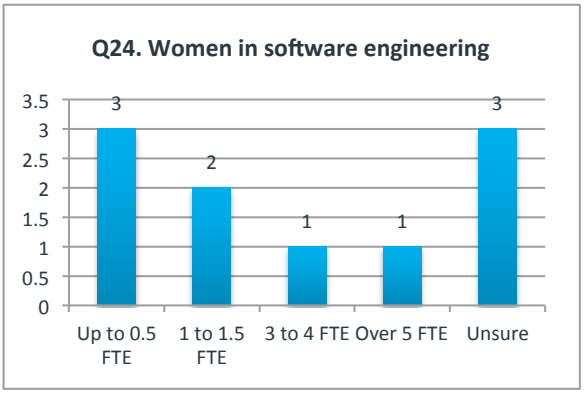


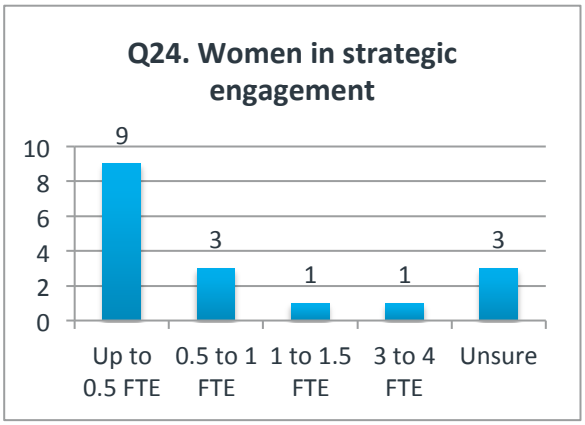
Q24 Women in HPC service management (FTE)

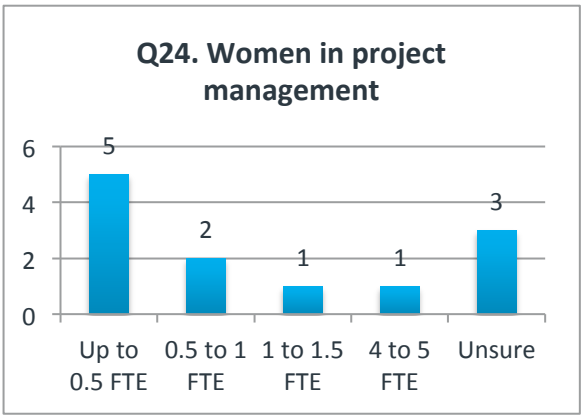


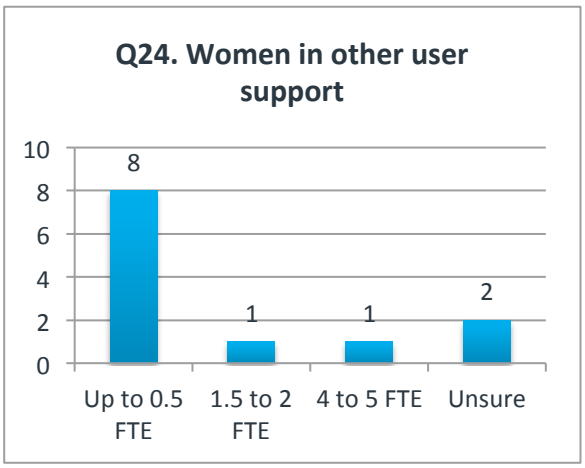












Training and User Support

Q25 What is your approach to training? Please tick all that apply

	HEI	Large / Specialist	Regional	Grand Total
We repurpose existing training materials developed elsewhere	12	5	2	19
We develop our own training	9	4	2	15
We run our own introductory training courses	10	3	2	15
We promote training available through e.g. Archer CSE and application vendors	8	2	1	11
We run our own advanced training courses	6	4	0	10

We provide training as part of a Centre for Doctoral Training	3	1	1	5
We provide training as part of a Doctoral Training Programme	2	1	1	4
Other	0	1	1	2
We do not provide training	0	0	0	0

Q26 URL for further information on training if applicable

- **https://virgodb.cosma.dur.ac.uk**
- **http://www.dirac.ac.uk/training.html**
- **https://www.epcc.ed.ac.uk/education-training , http://archer-www.epcc.ed.ac.uk/training/**
- **http://www.hpcwales.co.uk/solutions/skills-and-training**
- **http://www.shef.ac.uk/cics/research/training**
- **http://www.cardiff.ac.uk/arcca/services/events/index.html**
- **http://www.kcl.ac.uk/hpc/services/training.aspx**
- **http://www.ucl.ac.uk/isd/services/research-it/training**
- **https://www.acrc.bris.ac.uk/acrc/training.htm**
- **https://www.wiki.ed.ac.uk/display/ecdfwiki/Courses+and+Events**
- **http://wiki.rac.manchester.ac.uk/community/Courses**
- **http://www.arc.ox.ac.uk/content/training**
- We run our own introductory training courses
- **http://www.sussex.ac.uk/its**

Q27 How would you prefer to pool training across the sector?

	HEI	Large / Specialist	Regional	Grand Total
Share materials online freely and deliver locally	21	8	1	30
Marketplace of training materials and course delivery providers	8	4	2	14
Other...	1	1	1	3
Do not feel this would be beneficial	1	2	0	3

Q28 What proportion of your researchers are in self-supporting research groups?

	HEI	Large / Specialist	Regional	Grand Total
None or not known	6	1		7
<=25%	4			4
25% to 50%	6	2	2	10
50% to 75%	3	1	1	5
>75%	6	7		13
Grand Total	25	11	3	39

Q29 What are the major training challenges and which do you address locally?

	HEI	Large / Specialist	Regional	Grand Total
General introduction to programming	2	4	2	4
Targetted programming advice, e.g. using MPI, PE	3	4	2	4
Data science	4	4	2	4
Other..	4	3	1	4
Linux orientation	1	3	1	3
Application specific advice	3	3	2	3

Q30 Other training challenges

- Code optimization and porting to new technologies.
- As well as addressing local training challenges we aim to spread training expertise across the UK and beyond. All materials are CC licensed, we engage with other centres to help train their trainers, we develop online training

materials that can accessed from anywhere at anytime.

- Training the trainers i.e. enabling scientific communities to support themselves.
- Bioinformatics and genomics skills for scientists. Resourcing pressure has made this difficult to provide locally.
- Continual challenge as to the best ROI - tension between targeting the growth of new user communities (resource intensive) and focusing on the established researchers and enhancing their output (publications etc.) - the latter requires domain expertise by the trainers.
- Whether to implement charging tensioned against uptake by initial researcher interest How to best monitor the impact of the training through subsequent follow-up. Multi-stage questionnaire / survey - immediately post training and 3 months after.
- Training biologists to use Galaxy.
- Introduction to the HPC cluster - addressing scheduler, storage and best practices
- Helping users cope with the changes in workflow in moving from desktop computing to a more batch oriented; central HPC system, including the progression from GUIs to a more command line environment. This can take a good deal of effort with some new users.
- Delivering what looks like bespoke training to CDTs is a challenge. As is coordinating regional and national efforts to share resources. In reality we provide a broad range of training from many different sources. We also embrace training delivered by vendors to our researchers.
- Information about available software (incl. visualisation and analysis). Example batch jobs. Occasionally face-to-face training.

Q31 Other forms of support

	HEI	Large / Specialist	Regional	Grand Total
Informal networking for existing HPC users	17	7	3	27
HPC community wiki/forum	12	10	2	24
Seminar series featuring HPC based research	9	4	2	15
Speak/exhibit at researcher focussed events	10	3	2	15
Embed HPC staff in research groups	7	7	0	14
Direct work on research projects by HPC staff	8	6	0	14
Other.	2	3	0	5

Q32 Are you able to provide HPC case studies for (e.g.) HPC-SIG website, RCUK and Innovate UK e-Infrastructure SIG?

	HEI	Large / Specialist	Regional	Grand Total
No	16	5		21
Yes	6	6	3	15
Grand Total	22	11	3	36

Cloud and Shared Services

Q33 Which statements best characterise your organisation's use of "cloud" services?

	HEI	Large / Specialist	Regional	Grand Total
Some of our researchers use cloud services in an unsupported way	14	4	1	19
We don't use cloud services at all	6	2	2	10
Some of our researchers use cloud services with central support	4	3	0	7
Other....	2	3	1	6
We have a policy on use of cloud facilities	0	1	0	1
There are other cloud providers we would like to see the sector partner with	1	0	0	1
We use Jisc's cloud deals, e.g. Amazon portal	1	0	0	1
Most of our researchers use cloud services with central support	0	0	0	0

We exclusively use cloud based compute facilities	0	0	0	0

Q34 Use of cloud services - further information. Please describe the cloud services your organisation uses, or wishes to use.

- We use github for software version control and promote its (or equivalent alternative service) to our users. Many users use cloud services on their own, too, e.g. dropbox etc to share files but this is totally unsupported and not recommended.
- Dirac facilities are provided "as a service" from the point of view of the end users. Although this is not us using HPC as a Service, it is us providing HPC as a Service.
- EMBL-EBI Embassy Cloud - VMware & OpenStack
- **http://www.fortissimo-project.eu/** Web services for big data: EUDAT Investigating use of StarCluster for specific HPC workloads in the cloud
- Occasional Amazon S3 at small scale. VMs for ad-hoc web servers (short lived). Hosted VMs for delivering off-site training.
- We provide private cloud to our user base using a variety of technologies such as VCloud, Open Nebula
- Large scale use of Open Stack for research. VMWare used to support business activities.
- AWS at the moment. We'd like to be using OpenStack or similar ourselves, to provide an AWS-like experience to the users.
- We have made a number of efforts to launch our own Cloud services together with Fujitsu. This requires an enhanced HPC environment (security, ease of use with domain specific portals, billing and SLA contractual expertise) that are not typically not required by an academic
- We have evaluated AWS (including extending our cluster using Bright cluster manager), but found it unsuitable for many HPC applications.
- Some ad-hoc usage of Amazon cloud via Computer Science researcher groups. Undertaking a cost and benchmarking performance analysis in collaboration with BioScience in Q3/2015 Annual cost performance analysis of cloud vs. in-house provision presented to Governance Group
- We would be interested in some form of cloud-bursting for compute - or possibly SaaS for some products which would relieve the main HPC system of these smaller users.
- We are also a specialist HPC cloud service in our own right.
- Some researches (e.g. computer scientists) use IaaS cloud services in an unsupported way. We have also interfaced our HPC to Amazon AWS for "cloud bursting" experimentally but have not found the performance/price ratio to be beneficial at this time for raw HPC.
- Office 365, Azure, Simplicity
- Amazon Glacier. Arkivum.
- We use cloud for things like email & some working data store. We're in conversation with Microsoft about opportunities in scalar/throughput processing.
- We are interested in exploring what additional features of our existing services, or additional services, cloud resources would enable.
- We are investigating the ease with which users can potentially exploit the Amazon portal via the Jisc deal.
- Amazon, MS Azure etc. The use of these services is mostly unsupported but we have an approach that gives support to research no matter what platform they use. Cloud is part of that.

- Use of Amazon, MS Azure - especially where free access has been granted to support projects. In-house 'cloud' type services used for delivery of some HPC related computations - mainly from VMWare based systems
- Google Drive
- Developing private cloud and cloud bursting.
- Local / private cloud; we do not use public clouds (at least not in a centrally-supported way)

Q35 Which statement best characterises your organisation's approach to HPC as a shared service?

	HEI	Large / Specialist	Regional	Grand Total
We provide access to HPC as a shared service, e.g. EPSRC regional HPC centres	7	9	3	19
We may be interested in exploring shared services in the future	6	1		7
Not applicable	3	1		4
We use shared services provided for us by others	4			4
Other	3			3

Q36 Which areas of HPC provision might you share?

	HEI	Large / Specialist	Regional	Grand Total
Training	12	4	3	19
Application support	10	5	3	18
System management	10	6	0	16
End-to-end service provision	7	5	2	14
Data centre	5	7	1	13
Software development assistance	8	4	1	13

Other user support	7	4	1	12
Other	1	1	0	2

Research data management

Q37 Do you have a Research Data Management policy?

	HEI	Large / Specialist	Regional	Grand Total
No	2	1		3
Unsure	2	3		5
Yes	21	7	3	31
Grand Total	25	11	3	39

Q38 Does your organisation have a Research Data Management service?

	HEI	Large / Specialist	Regional	Grand Total
Not yet	6	6		12
Pilot service	9		1	10
Production service	6	3	2	11
Unsure	3	2		5
Grand Total	24	11	3	38

Q39 URL for further information about Research Data Management at your organisation if applicable

- **http://www.lib.cam.ac.uk/dataman/**
- **https://www.epcc.ed.ac.uk/facilities/uk-research-data-facility**
- **https://www.stfc.ac.uk/1930.aspx**
- At the moment we adopt RDM practices from our stakeholder institutions.
- **http://www.strath.ac.uk/researchdataproject/**
- **http://www.sheffield.ac.uk/library/rdm**

- **http://www.nottingham.ac.uk/research/research-data-management/index.aspx**
- **http://www.cardiff.ac.uk/insrv/researchdata/managingdata/index.html** (Converis-based solution)
- **http://www.kcl.ac.uk/library/researchsupport/research-data-management/index.aspx**
- **http://www.lancaster.ac.uk/library/rdm/**
- **http://www.lboro.ac.uk/service/research/offcampus/rdm.htm**
- **https://www.qub.ac.uk/directorates/ResearchEnterprise/ResearchPolicy/ResearchDataManagementPolicy/**
- **http://www.bath.ac.uk/research/data/**
- **http://data.bris.ac.uk/**
- **http://www.ed.ac.uk/schools-departments/information-services/research-support/data-management**
- **http://www.liv.ac.uk/csd/research-data-management/**
- **http://tinyurl.com/owpwj2o**
- **http://researchdata.ox.ac.uk/**
- **http://www.st-andrews.ac.uk/itsupport/academic/research/about/strategy/**
- **http://www.sussex.ac.uk/library/research/researchdatamanagement/**

Q40 What technologies have you deployed that can be regarded as being for Data Exploration ("Big Data") activities?

- Large shared memory computers; 100-terabyte-scale file-systems
- HTE, Spark/Hadoop
- DIR Machine (Data Intensive VM Cluster) RDF Cluster (Linux Containers based cluster for Data Analytics) SPRINT (MPI Parallel R, **http://www.r-sprint.org/**) SGI UV20
- 00 (Digital Health Institute, for complex data analytics
- IBM products: Big Insights, Streams, Infosphere Data Explorer, SPSS, Infosphere Content Analytics, Cognos, Watson
- SCD manages large scale scientific data in order of 50PB which is accessible at high data rates using a range of technologies appropriate to the customer. e.g. for
- JASMIN using Panasas with a multi terabit IO capacity. GridPP Castor using commodity disk solutions to SL85000 tap
- Tiered storage, looking at object storage for capacity in the next iteration.
- Tableau. R for Genome-Wide Association Studies (if those count as "Big Data")
- Hadoop, Cassandra, hive, pig, spark, maven, NoSQL databases.
- A file store dedicated to research data, a base level service is free of charge. Research groups can purchase additional storage. The facility links to the HPC facility.
- Preparing infrastructure (storage, networks) to support future 'Big Data' research.
- Galaxy web service, Hadoop, NoSQL databases.
- Nothing at present. Will be looking into Hadoop in the near future
- We have funded a small hadoop cluster as a test bed. We're looking at image library technologies (mainly for medical research projects)
- Hadoop
- GPFS, Aridhia, large memory systems
- pbdR supported on cluster; site license (via Engineering) for TecPlot visualisation software as well as several open source data exploration and visualisation tools.
- We have specifically deployed a large shared memory facility in support of analysis of public health records (Farr Institute facility as part of N8 HPC) and we have a data analysis cluster delayed institutionally.
- none directly yet by ARC. Others within the university explore different technologies such as hadoop

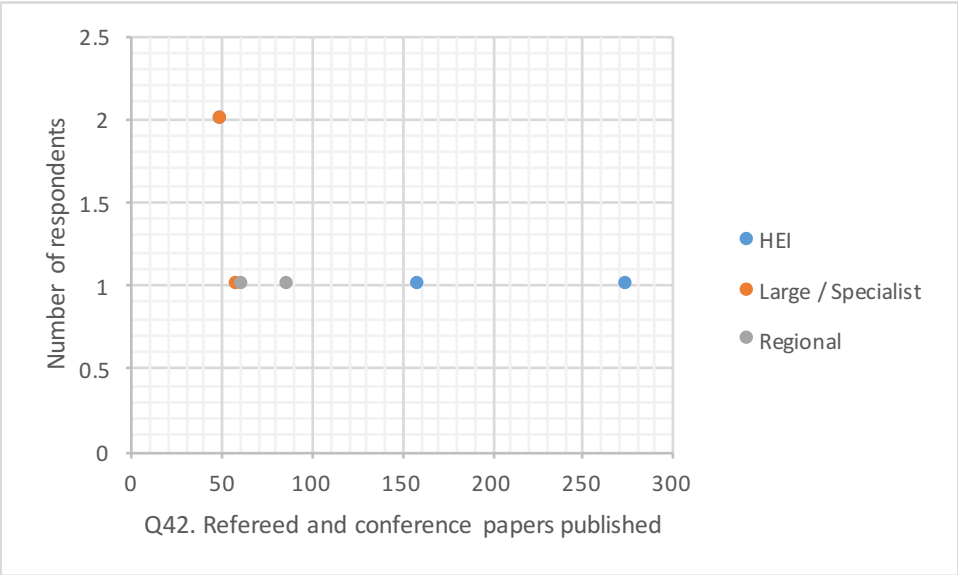
- Small 7 node mapr installation.
- Installed R for chem/bioinformatics
- SQL services
- Data analytics cluster (Hadoop; streaming analytics); local cloud for bioinformatics

Academic impact

Q41 What are your Key Performance Indicators to measure academic impact ?

	HEI	Large / Specialist	Regional	Grand Total
Amount of research in terms of research grants	14	4	3	21
Number of different research areas using the systems	15	3	2	20
Number of post-grad and post-doctoral users	6	3	2	11
Number of projects run on systems	13	5	3	21
Number of research papers published	14	8	3	25
Number of theses produced	2	2	3	7
Number of users	18	5	3	26
Other.....	4	2	1	7
Research highlights - high profile papers, breakthroughs, news items	10	7	2	19

Q42 Refereed and conference papers published in the last year



Q43 URL listing these publications if applicable

- **http://www.cosmos.damtp.cam.ac.uk/info/cosmos-publications-2014/**
- **http://www.dirac.ac.uk/science_all.html**
- Separate report available on request.
- **https://pure.strath.ac.uk/portal/en/equipment/uoshpc(a35d23ad-d4fa-481c-b7db-e535add521f7).html**
- Contained in a separate report - available upon request
- See below report

Q44 Does your organisation produce an Annual Report? If so, please enter its URL below

- COSMOS annual reports are part of the DiRAC Annual Reports at **http://www.dirac.ac.uk/**
- **http://www.dirac.ac.uk/science_all.html**
- **http://www.ebi.ac.uk/about/brochures**
- Different reports for different projects and services
- **http://www.stfc.ac.uk/SCD/resources/PDF/SCD_Science_Highlights_2014.pdf**
- In progress
- **http://cics.dept.shef.ac.uk/reports/cics-annual-report-2013.pdf**
- **http://www.cf.ac.uk/arcca/news/annualreport2015.html** (in press)
- **http://www.kcl.ac.uk/newsevents/publications/report.aspx**
- Our annual report is an internal document and presented to our HPC Board. However, we do produce public reports usually to coincide with a new system, e.g.: **https://www.acrc.bris.ac.uk/acrc/HPC_report.pdf**
- **http://www.cam.ac.uk/annual-report**
- **https://www.liv.ac.uk/annual-report/**
- **http://n8hpc.org.uk/n8-hpc-annual-survey-2014-released/**

- **http://www.it.ox.ac.uk/about/reports/it-services-annual-report-20132014**

Q45 Does your organisation publish Research Highlights? If so can you provide URLs for up to 3 recent highlights

- **http://www.cosmos.damtp.cam.ac.uk/info/research-highlights-2013-14/**
- **http://www.dirac.ac.uk/science_news.html**
- Different for different projects and services
- For SCARF: **http://www.scarf.rl.ac.uk/sites/default/files/docs/RAL-TR-2014-017.pdf**
- **https://www.sanger.ac.uk**
- Yes - to appear on new web site that is currently under construction - available on request
- **http://www.archie-west.ac.uk**
- **http://hpchub.sites.sheffield.ac.uk/research-groups**
- **https://www.nottingham.ac.uk/research/news.aspx**
- Yes - contained in the annual report
- **http://www.ucl.ac.uk/research-it-services/rits-case-studies** more in development at present
- Yes, but these are usually published by our PRO, for example: **http://www.bristol.ac.uk/news/2014/november/hendra-in-bats-and-humans.html**
- **https://www.liv.ac.uk/research/news/**
- In a manner of speaking, again see N8 HPC website and annual report.
- **http://www.sussex.ac.uk/research/**

Who uses your service?

Q46 Do you provide HPC services beyond your immediate organisation?

	HEI	Large / Specialist	Regional	Grand Total
HEIs	3	3	1	7
Other.....	3	3	1	7
DiRAC	2	4		6
GridPP	3	2		5
N8 HPC	4	1		5
Biochemistry and Life Sciences Informatics	1	3		4
ARCHER	3	1		4
Farr Institute	3	1		4

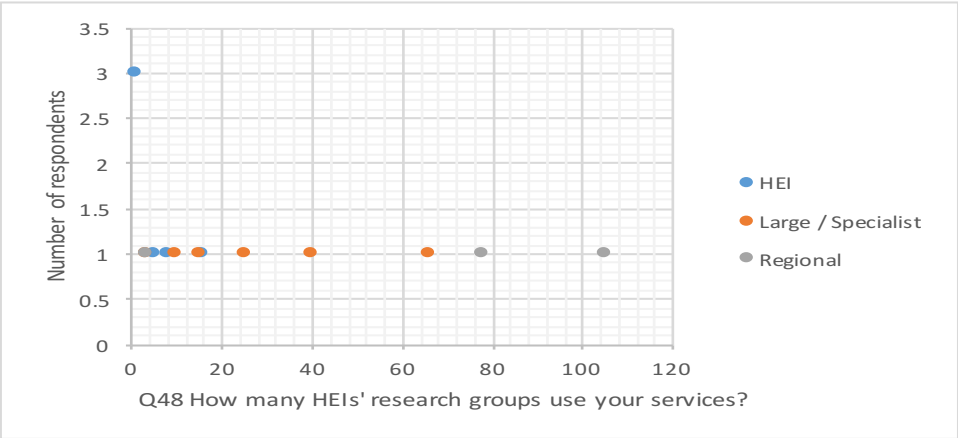
SES/CFI	2	1	3
JASMIN/CEMS		2	2
Cancer Research UK	1	1	2
Genomics England	1	1	2
HPC Wales	1	1	2
National Oceanographic Centre	1	1	2
Research Data Facility	1	1	2
MidPlus	2		2
ARCHIE-WeSt		1	1
Diamond Light Source		1	1
ELIXIR		1	1
HPC Midlands		1	1
Hartree Centre		1	1
Medical Bioinformatics Initiative		1	1
Wellcome Trust Sanger Institute		1	1
Administrative Data Research Centres			
The Genome Analysis Centre			

Q47 What services do you provide to third parties?

	HEI	Large / Specialist	Regional	Grand Total
Compute and storage	8	10	3	21
Support and training	3	6	3	12
Data services		7	2	9

Networking		7	2	9
Software development	1	4	1	6
Business development	1	2	2	5
Other.....	1			1

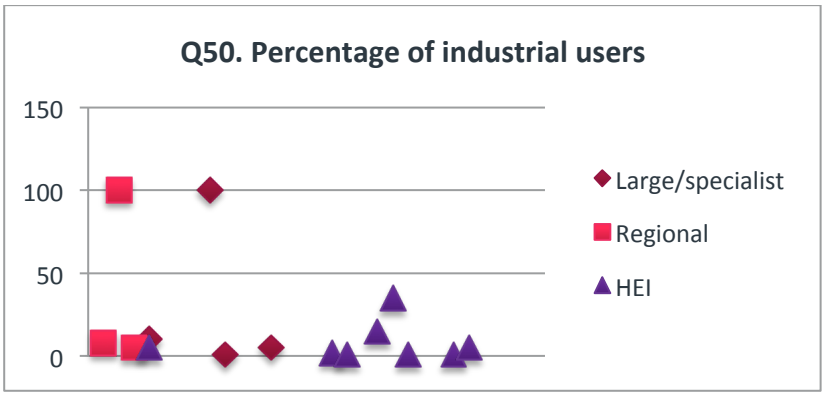
Q48 How many HEIs research groups use your services?



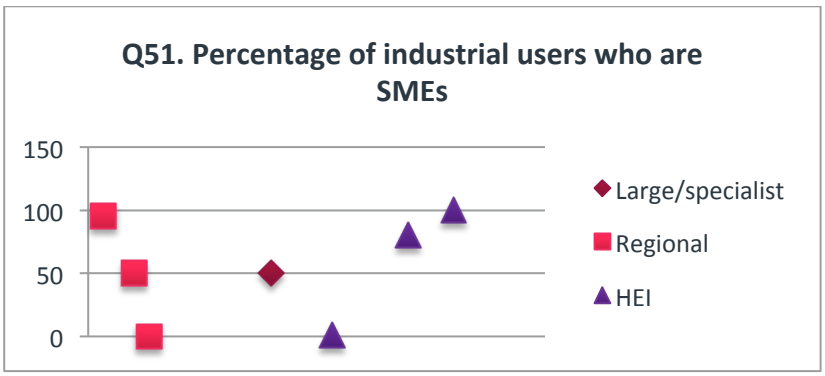
Q49 Public and third sector organisations

	HEI	Large / Specialist	Regional	Grand Total
NHS and/or Department of Health	3	6	2	11
Charitable sector		6	1	7
Department for Business, Innovation and Skills	1	5		6
European Commission		2		2
European Space Agency		1	1	2
Other.....	1	1		2
Department of Energy and Climate Change				

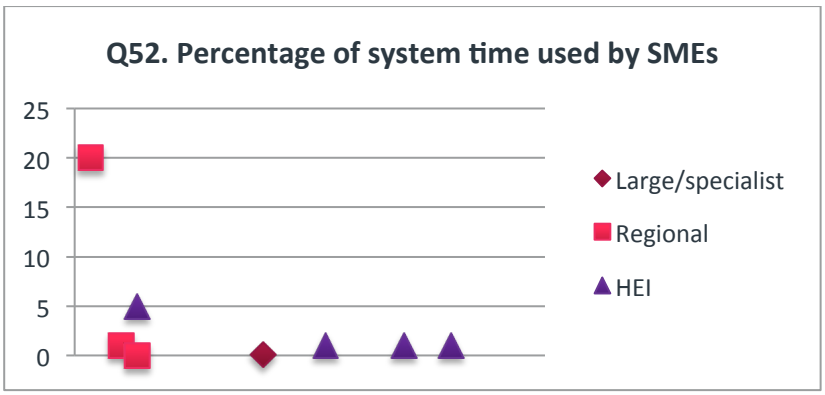
Q50 What percentage of your users are from industry?



Q51 Of industry users, what proportion are SMEs?



Q52 What percentage of system time is being used by SMEs? Percentage of system compute resource over a twelve month period



Q53 Are you taking any specific steps to increase SME uptake?

- Yes; advertisement and personal approach.
- Industry programme

- Have always been on the forefront of engaging SMEs to use HPC and continue to work extensively in this area.
- Yes. This is the Hartree Centre mission. We have a Business development team of 5 people devoted to this.
- The Hartree Centre is driving SME uptake of HPC and will provide a separate response to this survey.
- No, focus in the sector is more on larger Pharma.
- Institute is planning a Biodata Innovation Centre for up to 30 SME's
- HPC Wales Stage 1 is currently coming to a close, with the direction of travel for its successor project yet to be quantified. While SME engagement will remain of importance it is likely, based on initial discussions with WEFO and the WG, that the profile will broaden
- Yes, dedicated business development manager looking to work with SMEs
- Holding knowledge exchange events and exhibiting at industry events such as NAFEMS. We have also undertaken a security audit of our facility by an ISO compliant organisation. We are also in the process of attempting to secure internal funds for a full time Business Development role
- Work with business engagement teams, the Advanced Manufacturing Research Centre and Advanced Computing Research Centre.
- As a partner in HPC Wales, we are developing our industrial contacts through this mechanism.
- The CORE initiative with Imperial College brings these people in from time to time.
- Our strategy for SME involvement mirrors our embedded use of institutional BE teams which we coordinate in order to support research that uses HPC to engage with industry. Broadly, our approach is to target supply chains of larger companies.
- Increasing engagement with University consulting services to broaden outreach of HPC activity.
- Liaison with Sussex Innovation Centre

Q54 What sectors are being served by your industrial users?

	HEI	Large / Specialist	Regional	Grand Total
Advanced materials and manufacturing	7	1	3	11
Energy and environment	6	1	3	10
Life sciences	3	4	2	9
Finance and professional services	3	1	2	6
Defence and security	4	1	1	6
Transport	3	1	1	5
Creative industries	1	1	2	4
Construction	2	1	1	4
Other.....		1	1	2

Further information

Q55 Please use this space for any further information you would like to provide

- As with the system survey, this has proved difficult to complete from the perspective of a distributed organisation that is providing HPC services to both academic and third party organisations. 2. The ability to save and restart the survey would have been helpful.

- Software licensing still a major issue for SME's, both in terms of providing a barrier to access and in terms of acting as a constraint on the extent to which they can leverage the power of HPC. Consequently, the gains we are making with SME's are not reflected in the report.

- Given the length of the survey it would have been helpful to have a "save now and complete later option" - plus being able to print or make copies of the completed survey report for our own records (helps future responses!)

- The survey questions are often unclear.

- There is active industrial work going on with the separate VEC (Virtual Engineering Centre) in partnership with Hartree. All industrial use of Liverpool HPC facilities are via collaborating academics.

Appendix G – Hardware: The survey questions

Q1 Organization name
Q2 Organizational unit
Q3 Email address
Q4 Job title
Q5 System name
Q6 External IP address of FQDN
Q7 URL for the website of the system or overall service
Q8 What are the top three research areas the system is used for?

Hardware specifications

Q9 Total number of processor cores in the system
Q10 Number of compute nodes
Q11 Number of processor cores per compute node
Q12 RAM per core (Gigabytes)
Q13 Compute node processor specification e.g. Intel Ivy Bridge E5-2670v2 2.5GHz
Q14 How many GPU equipped nodes does the system have?
Q15 How many Xeon Phi equipped nodes does the system have?
Q16 How many "fat" nodes does the system have, i.e. >=8GB RAM per core
Q17 Does the system have a dedicated Visualization capability? Please describe below if applicable
Q18 Interconnect Switch Fabric e.g. QDR/FDR InfiniBand, Gigabit Ethernet, NUMAlink
Q19 When was the system commissioned?
Q20 When will maintenance for the system terminate?

Storage

Q21 Describe the storage component of the system e.g. NetApp CDE5400, Panasas ActivStor 11, Direct Attached Storage on compute nodes
Q22 Total usable storage for HPC users
Q23 What file system(s) and/or object stores do you use for shared storage?
Q24 Do you split system storage in terms of TB between fast, tertiary, archive storage?
Q25 Number of registered users

Performance and connectivity

Q26 Theoretical Peak Performance (Tflop/s)
Q27 Node to Node Data Rate (Gbit/s)
Q28 Average node to node latency (Microseconds)

Q29 Typical CPU load as a % of overall system
Q30 Peak inbound sustained data transfer rates
Q31 Peak outbound sustained data transfer rates
Q32 Is the bandwidth above dedicated for HPC service use?
Q33 Special connectivity requirements e.g. Lightpath, dedicated circuit, low latency

Software and operating environment

Q34 Installed software
Q35 What is the primary Operating System you use on compute nodes?
Q36 What is the primary Operating System you use on head/login nodes?
Q37 What scheduler(s) do you use?
Q38 Do you provide a Web Portal to your users? If so, please describe below
Q39 Do you back up HPC user data?
Q40 Do you do have scheduled maintenance and if so how often?

Access, authorization, accounting and identities

Q41 Management of HPC system users and projects
Q42 Accounting and resource allocation

Further information

Q43 Please use this space if there is any other information you would like share about the system

Appendix H – Hardware: Summary of the survey data

Science areas - of the 75 systems that respondents reported on, key application areas were Computational Fluid Dynamics (24%) materials science (16%) life sciences and bioinformatics (37%) and physics (41%).

Hardware - with the growth in processor core counts, many institutional systems were now on a bar (in core count, but not necessarily in capacity) with a number of the regional and large/specialist facilities. Several HEIs also had more compute nodes than the regional centres and a number of the large/specialist facilities. This could be expected to change with the next round of hardware refresh. 39% of the systems in the survey now had GPU equipped nodes, although in most cases these were 20 nodes or less out of a whole cluster. 19% of the systems in the survey had Xeon Phi equipped nodes, but virtually all of these were 4 or less nodes out of a cluster.

Fat (high memory) nodes - the vast majority of HEI and regional systems still only have a handful of "fat" nodes, with 8GB or more of memory per core. Systems with hundreds of high memory nodes are still the domain of the large and specialist facilities.

Visualization - 37% of respondents' systems had a dedicated visualization capability.

Interconnect - 48% of respondents' systems used QDR Infiniband, with a grand total of 65% of all systems in the survey being connected via Infiniband. 23% of systems in the survey were connected with Gigabit Ethernet, with 13% operating at 10Gbit/s, and 2 systems operating at 40Gbit/s.

Lifecycle - many respondents were still running off-maintenance systems commissioned as far back as 2007 and 2009. 8 new systems had been commissioned in 2014, and a further 6 in 2015. 26 systems (35% of the UK NeI) would fall off maintenance in 2015 and 2016.

Storage - institutions are now starting to inch towards the sort of Petabyte scale storage solutions which had been very much the domain of large and specialist facilities, with 5 HEIs having storage facilities of 1PB or above for their scientific computing operation. Lustre (33%), NFS (25%) and GPFS (18%) are still by far the most common filesystems, although growth in Panasas installations has seen PanFS rise to 14% of survey respondents' systems.

User base - 28% of the systems in the survey have less than 100 users, with only 16% having more than 1000 users.

Performance - as with core counts, institutional systems have started to catch up with tier 1 and 2 facilities, with 21 out of 75 (28%) of systems in the survey running at 100 Teraflops/s or better. However, not all respondents provided performance stats for their systems.

Utilization - 51% of systems were running at 75% load or above, including oversubscription, with 75% generating peak inbound data rates of between 1 and 10 Gbit/s. 19% of the systems in the survey were generating over 10Gbit/s outbound traffic. 30% of systems had been provided with dedicated bandwidth, either in their own right, or shared with other

equipment.

Operating environment - 59% of systems used an unencumbered Linux distribution (such as Debian or CentOS) as their primary compute node operating system, with 49% also using an unencumbered distribution for their login nodes. By contrast, fully supported distributions such as RedHat Enterprise Linux (RHEL) were found on 24% of login nodes and 15% of compute nodes. Only one respondent was running Windows HPC Server.

Scheduler - Sun Grid Engine or its derivatives continue to dominate the schedulers in the survey responses, running on 24% of systems. However, SLURM has now grown to 16% of the installed base, with 12 installations across the UK NeI.

Backups - 32% of systems are not backed up at all, versus 30% that are backed up, and 37% where some data (not necessarily including user files) is backed up.

Scheduled maintenance - 35% of systems in the survey do not have scheduled maintenance. 43% of systems have either monthly, quarterly or annual scheduled maintenance.

Management of users and projects - 41% of UK NeI systems have authentication linked to institutional systems such as Active Directory. However, 44% of systems have at least some manually curated accounts. 32% of systems have a peer review process for new projects, although just 19% have an equivalent process for new user accounts. Only 9% of respondents automatically delete expired user accounts, and just 11% update user accounts automatically to take account of changes such as moving departments. 13% of respondents were interested in trialling Jisc's Assent service (formerly Project Moonshot).

Accounting and resource allocation - 51% of respondents reported that projects on their systems were given a resource allocation, however 23% stated that they did not impose any resource limits on their users. 16% (12 institutions) reported that usage of their systems was often constrained by software licenses.

Appendix I – Hardware: Full break down of survey data

Q8 What are the top three research areas the system is used for?

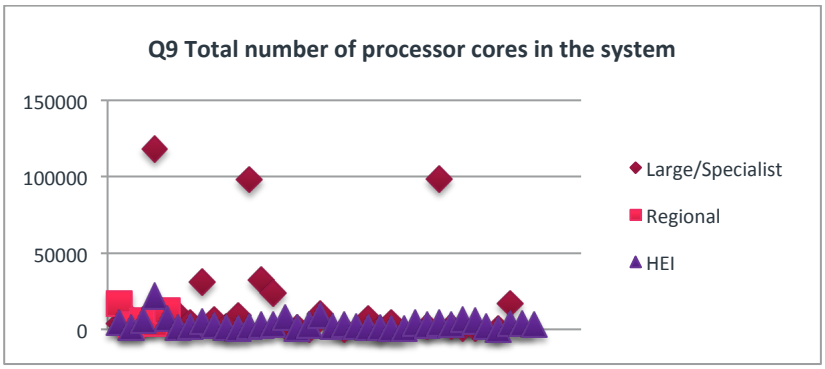
Top research areas	Number of responses
Adaptive Systems	1
Aeronautical Engineering	1
Aerospace Eng	1
Astronomy	3
Astrophysics	6
Atomic structure	2
Big Data and Data Analytics	2
Biochemistry	1
Bioinformatics	4
Biological Sciences	4
Biomedical Sciences	2
CFD	18
Cancer Research	1
Chemical Engineering	1
Chemistry	8
Civil and Environmental Engineering	2
Climate/Ocean Modelling	4

Computational Chemistry	4
Computational Medicine	1
Computational Neuroscience	1
Computational Physics	1
Computer Aided Formulation	2
Condensed Matter	2
Cosmology	5
Cryo-electron microscopy	1
Earth Science	4
Economics	1
Electronic System Design	1
Energy Efficient Computing	1
Energy Efficient Transport	1
Engineering	7
Environmental Genomics	1
Exoplanets	1
FEA	2
Fundamental Physics	1
Gait Analysis	1
Galaxy Formation	2
Genomics	4
Geographical Sciences	1
Healthcare	2
High Energy Physics	1
Hydrology	1
Informatics	2

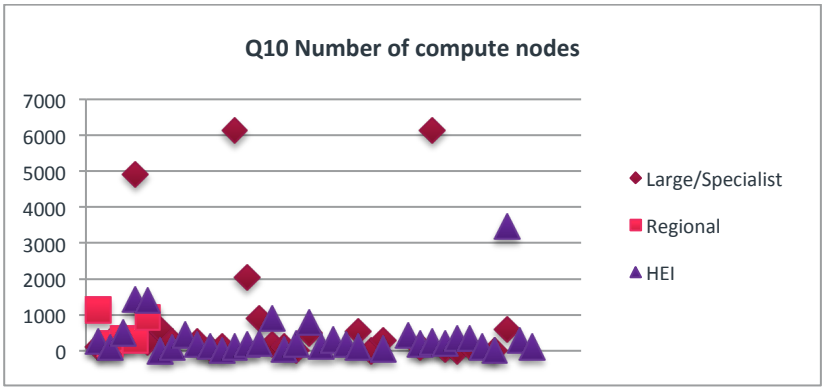
Life Sciences	6
MHD	2
Machine Learning	1
Materials Science	12
Mathematics	5
Mechanical Engineering	1
Medical Informatics	1
Microbial bioinformatics	1
Molecular Dynamics	3
Natural Language	1
Next generation sequencing	1
Optoelectronics	1
Particle Physics	5
Physics	6
Plasma Physics	3
Pyschology	1
QCD	1
RNA sequencing	1
Satellite images	1
Science	1
Semiconductor Device Modelling	2
Soft Matter Physics	1
Speech and image processing	1
Weather/climate	1

Hardware specifications

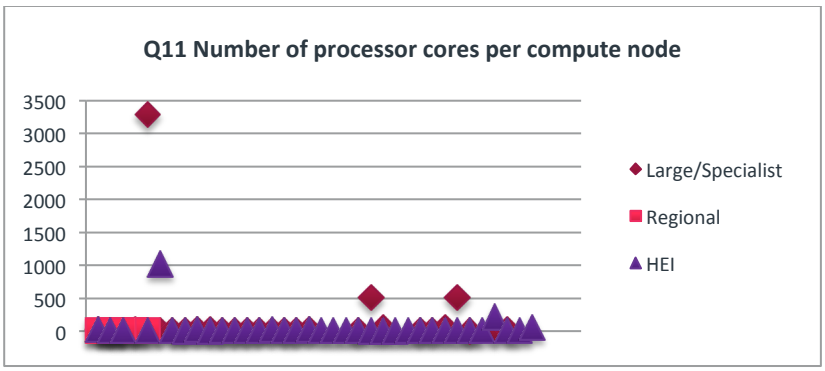
Q9 Total number of processor cores in the system



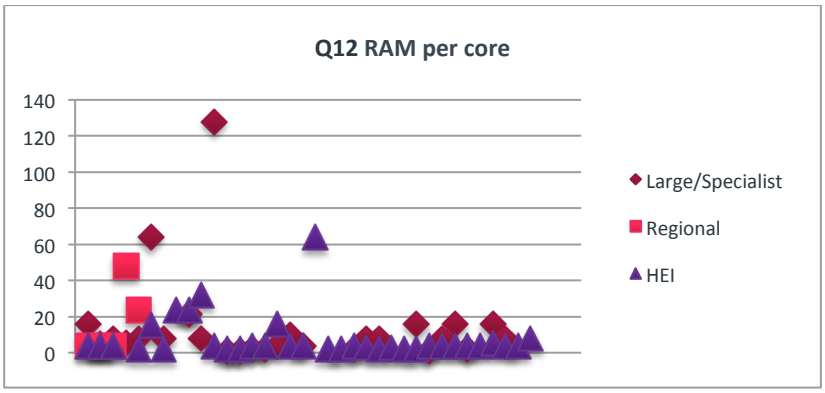
Q10 Number of compute nodes



Q11 Number of processor cores per compute node



Q12 RAM per core (Gigabytes)

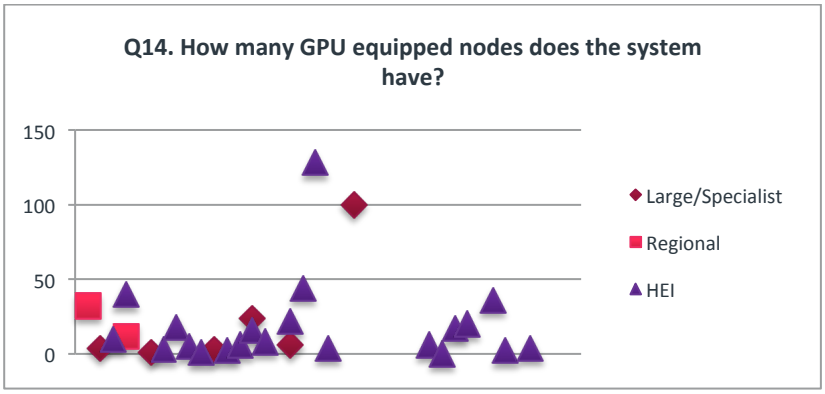


Q13 Compute node processor specification e.g. Intel Ivy Bridge E5-2670v2 2.5GHz

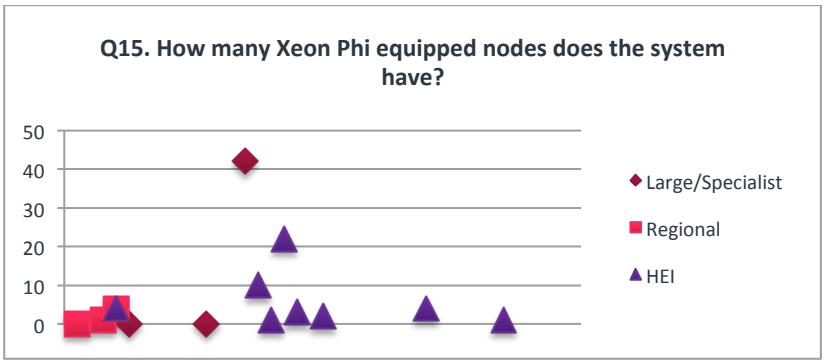
- 16 x E5530 2.40GHz 32 x X5650 2.67GHz 2 x AMD6174 2.2GHz 30 x E5-2650v2 2.6GHz 32 x AMD6378 2.4GHz
- 220*E5520, 156*E5-2650 v2
- AMD Opteron 6276 2.3 GHz
- AMD Opteron 6376 2.3GHz
- AMD Opteron Processor 6378 2.4 GHz
- E5-2580v2
- E5-4620 v2 2.6GHz
- E5462 2.8GHz, E5-2670 2.6GHz
- IBM Power
- IBM PowerPC A2 @ 1.6GHz
- Intel
- Intel Atom dual core processor @ 1.60GHz
- Intel E5 2650V2
- Intel E5-2670 2.60GHz
- Intel E5-4650L 2.6GHz 8 core Xeon processor
- Intel E5620
- Intel Haswell E5-2640v3 2.6GHz
- Intel Ivy Bridge E5-2650 2.6GHz
- Intel Ivy Bridge E5-2650 v2 2.6 GHz
- Intel Sandy Bridge E5-2640 2.50GHz
- Intel Sandy Bridge E5-2650
- Intel Sandy Bridge E5-2660
- Intel Sandy Bridge E5-2670 2.6GHz, Intel Westmere X5660 2.8GHz, Intel Haswell E5-2680 V3 2.5 GHz
- Intel SandyBridge E5-2660 @ 2.20GHz
- Intel Sandybridge E5-1650 3.2GHz
- Intel Sandybridge E5-2650 2.0GHz
- Intel Sandybridge E5-2670 2.6GHz
- Intel Westmere E5-2697 X5650 2.67 GHz; Intel Sandy Bridge E5-2690 2.9 GHz; Intel Sandy Bridge E5-2670 2.6 GHz;
- Intel Westmere X5650
- Intel Xeon E5-2660 2.2GHz
- Intel Xeon E5-2695v2
- Intel Xeon E5530 2.40GHz
- Intel Xeon X5660 2.80GHz

- Intel(R) Xeon(R) CPU E5-2630 v2 @ 2.60GHz
- Intel(R) Xeon(R) CPU E5-2660 0 @ 2.20GHz
- Intel(R) Xeon(R) CPU E5-2670 0 @ 2.60GHz
- Intel(R) Xeon(R) CPU E5-2670 0 @ 2.60GHz (SandyBridge)
- Intel(R) Xeon(R) CPU E5-2697 v2 @ 2.70GHz
- Intel(R) Xeon(R) CPU E5-4650L 0 @ 2.60GHz
- Intel(R) Xeon(R) CPU X5650 @ 2.67GHz
- Ivy Bridge UoB: E5-4620v2 2.6Ghz, E7-8857v2 3.0Ghz, E7-8880v2 2.5Ghz other sites: E5-4610v2 2.3Ghz, E7-8850-v2 2.3Ghz
- Ivybridge-v2 E5-2650v2 2.6GHz
- Mix of Woodcrest, Nehalem, Ivy Bridge (various models thereof)
- N/A
- NA
- Nehalem X5560 @2.8Ghz, Westmere X5650 @2.7Ghz
- SCARF has many generations of processor as the model is to upgrade it yearly
- Sandy Bridge,Westmere, Haswell, Opteron
- Westmere X5650 @ 2.67GHz
- Westmere 2.8/2.95GHz
- Xeon E5-4620 v2 (20M Cache, 2.60 GHz)
- Xeon(R) CPU E5-2670 2.60GHz (3328 cores) for "regular" compute node
- latest intel generation
- varies
- various generations
- various generations latest install 2015
- westmere/ivybridge

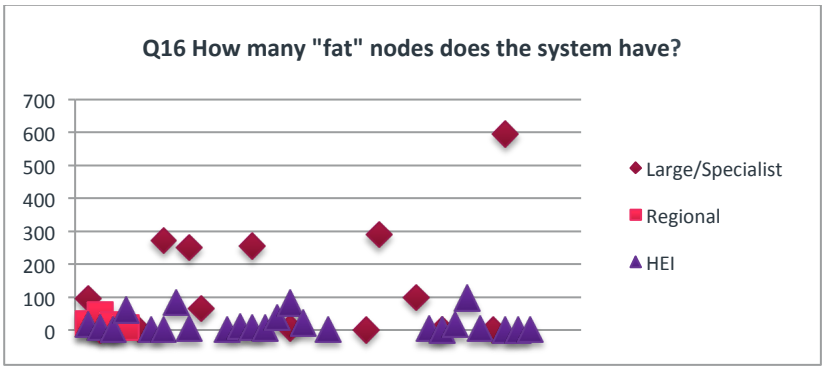
Q14 How many GPU equipped nodes does the system have?



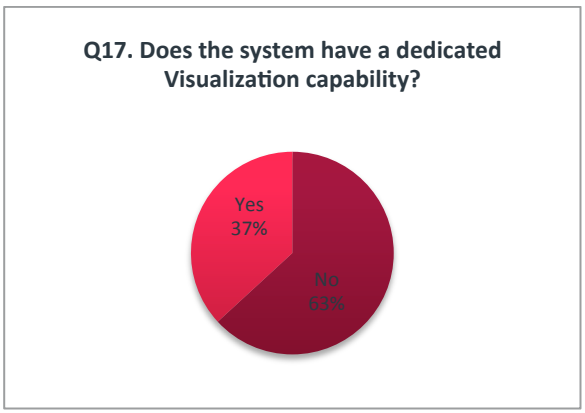
Q15 How many Xeon Phi equipped nodes does the system have?



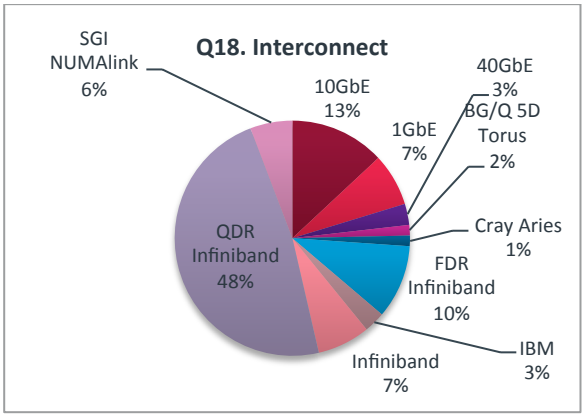
Q16 How many "fat" nodes does the system have, i.e. >=8GB RAM per core



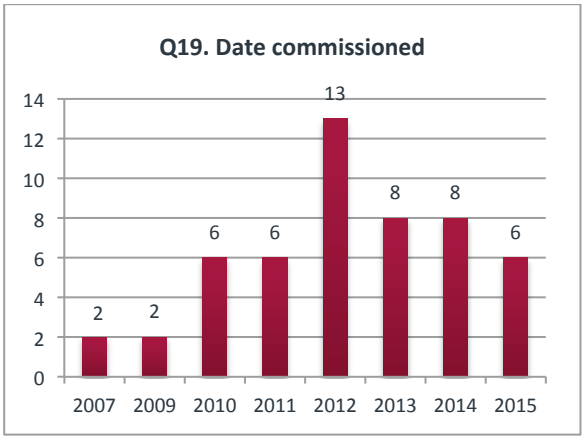
Q17 Does the system have a dedicated Visualization capability? Please describe below if applicable



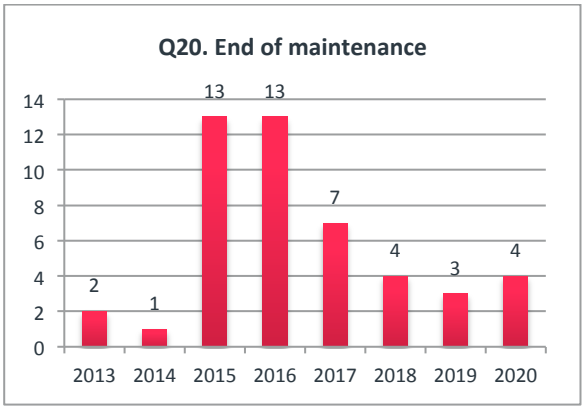
Q18 Interconnect Switch Fabric e.g. QDR/FDR InfiniBand, Gigabit Ethernet, NUMAlink



Q19 When was the system commissioned?



Q20 When will maintenance for the system terminate?



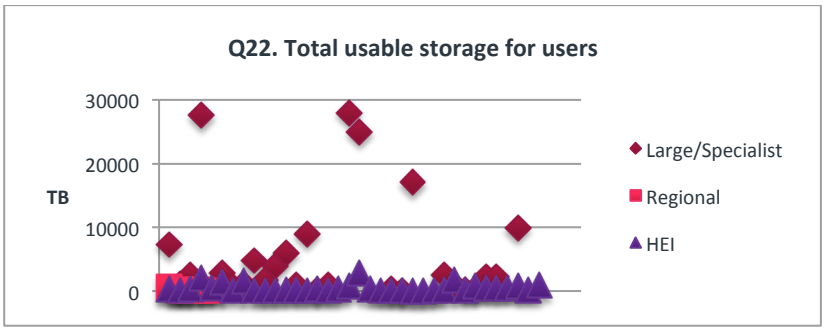
Storage

Q21 Describe the storage component of the system e.g. NetApp CDE5400, Panasas ActivStor 11, Direct Attached Storage on compute nodes

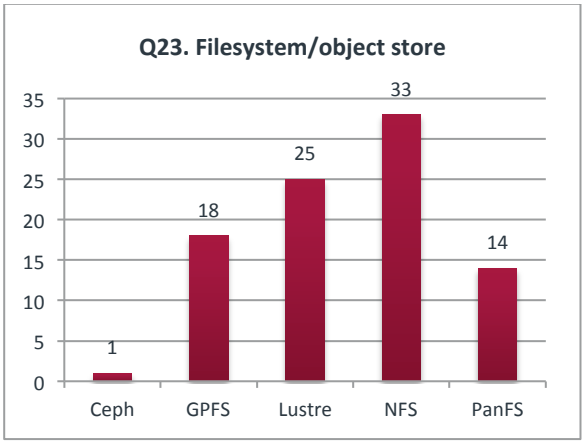
- 150TB Lustre
- 22 TB Scratch using SGI Storage Directly Attached 1x Panasas ActivStor 11
- 250TB lustre scratch 0.51PT Isilon resilient storage
- 2xSL8500 tape robots (1 used by Tier1) with 10000 slots each with a total nearline capability of 160PB (when fully populated) Tape Servers, Oracle Databases, Data Transfer Nodes supporting CASTOR and DMF front ended by a 700TB disk cache Monitoring Systems
- 7 shelves of Panasas ActiveStor 11
- 800TBytes Lustre Dell Power Vault 200 TBytes NFS Power Vault
- 8x HP P2000 G3 SAS disk trays
- Block device accessible via VM in user configured way
- Cloud: Isilon & SAN, Cluster: GPFS
- DDN
- DDN 9550, Panasas 14s
- DDN SFA 10K
- DDN SFA12K-40
- Dell MD storage.
- Dell MD3000
- Dell commodity
- Direct Attach on Server
- Direct Attached Storage
- Direct Attached Storage on compute nodes
- Direct Attached Storage on storage nodes
- Direct Attached storage through head node
- Direct attached 14PB of commodity storage using 400 disk servers. SL85000 tape robot with >24 drives with 14PB managed by CASTOR
- Direct attached storage
- Direct attached storage on nodes (0.5TB per node) DDN SFA10K/SFA12K HP SL4540
- EMC Isilon
- Each site has ~ 400TB GPFS attached to hypervisor nodes for VM images and working data. 6.9Pb of CEPH object store across the sites is being commissioned still
- Fast - Panasas ActiveStor 12 Tertiary - HNAS AMS2500 & HUS150
- Fraunhofer/BeeGFS parallel filesystem over IB using Dell MD3460
- GPFS
- GPFS Based upon DDN discs via Infiniband
- GPFS Based upon DDN discs via Infiniband. GPFS based upon IBM GSS24
- GPFS FPO based upon IBM GSS24
- GPFS across cluster with 250GB local storage per node
- GPFS based upon IBM GSS24 and GPFS based upon flash memory directly connected to I/O Nodes
- IBM DS3512, DS3524 and IBM DCS3700

- IBM DS5300
- IBM GSS24 + GSS26
- IBM general parallel file system (GPFS), giving around ~110TB of usable storage divided among the users and split between permanent storage and temporary/scratch space.
- LSI
- LSI SAS
- Long Term NAS: HP SL4540 based NFS storage totalling 696TB usable, split across two sites for disaster recovery High-speed scratch: Intel Enterprise Edition Lustre on Supermicro based hardware provided by Boston/BIOS-IT, 2 filesystems each with 512TB usable
- Lustre 120 TB
- Lustre NFS
- Lustre Parallel file system, NFS home file system,Nexenta
- Lustre scratch partition and backed up NFS home
- Netapp 176TB usable (for data and home areas) and lustre 260TB (fastdata area with 90day lifetime policy)
- PanasFS and FhGFS
- Panasas ActivStor 11
- Panasas ActivStor 11 GPFS
- Panasas ActivStor 12
- Panasas ActivStor 14
- Panasas ActiveStor
- Panasas ActiveStor 12
- Panasas ActiveStor 14
- Panasas ActiveStor 8 (being upgraded to 16 this year) for general use Multiple NFS disk storage nodes for High Energy Physics
- RAID 5 array within headnode
- SGI CXFS
- SGI IS5000
- SGI InfiniteStorage 5000 RAID Array IS5000 dual controller 2U 24-bay 2.5” RAID array containing: o 20 x 3000GB 7.2K rpm 2.5” 6Gb/s SAS HDD
- Storage array comprising Dell MD3200 and MD1200 units with RAID 6 arrays
- Storage is NFS mounted from a single storage node.
- Sun/Oracle Snowbird system made up of 8 J4440 arrays
- Two NetApp E2600
- Various Dell MD storage.

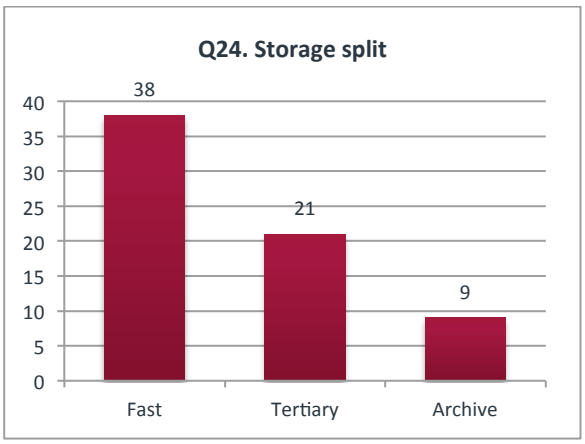
Q22 Total usable storage for HPC users



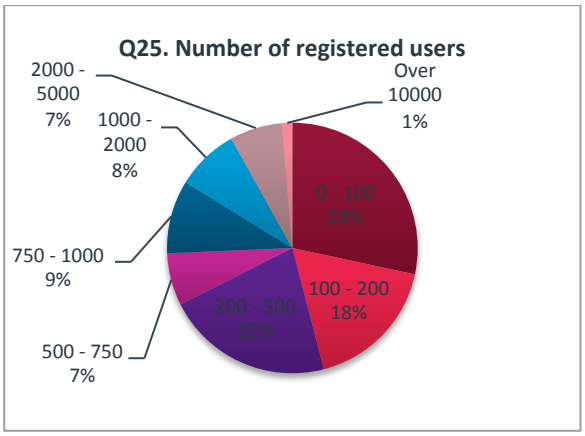
Q23 What file system(s) and/or object stores do you use for shared storage?



Q24 Do you split system storage in terms of TB between fast, tertiary, archive storage?

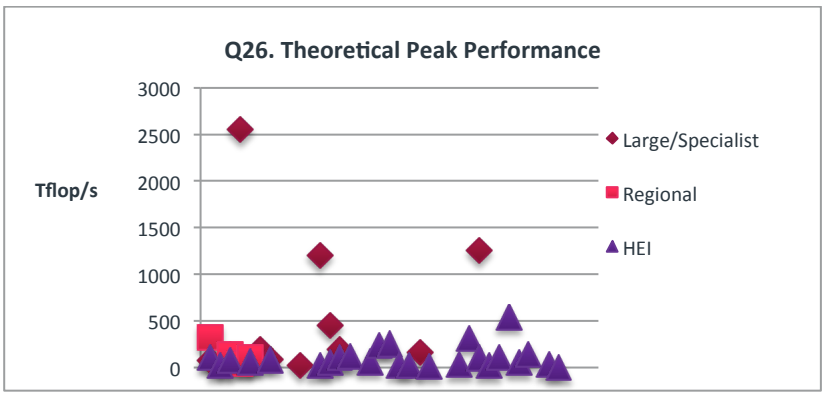


Q25 Number of registered users



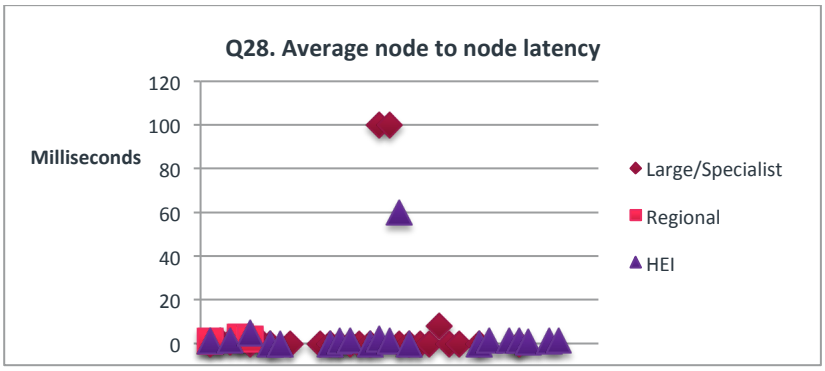
Performance and connectivity

Q26 Theoretical Peak Performance (Tflop/s)

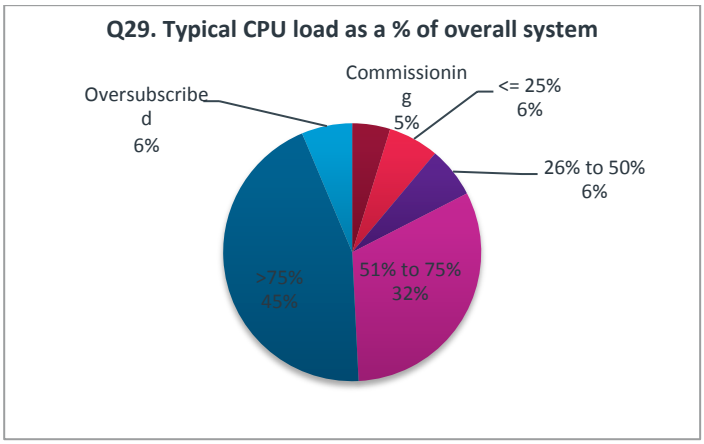


Q27 Node to Node Data Rate (Gbit/s)

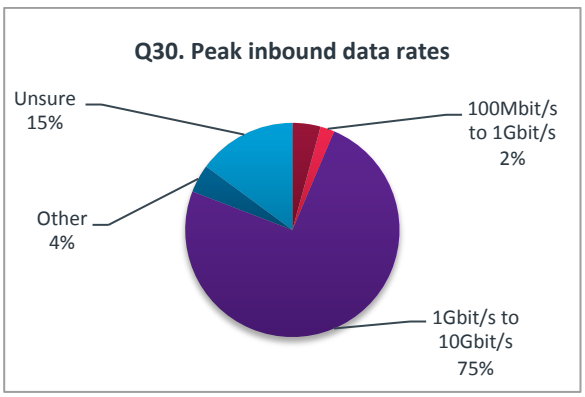
Q28 Average node to node latency (Microseconds)



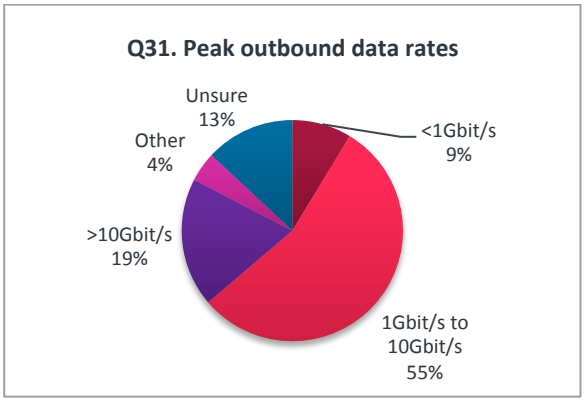
Q29 Typical CPU load as a % of overall system



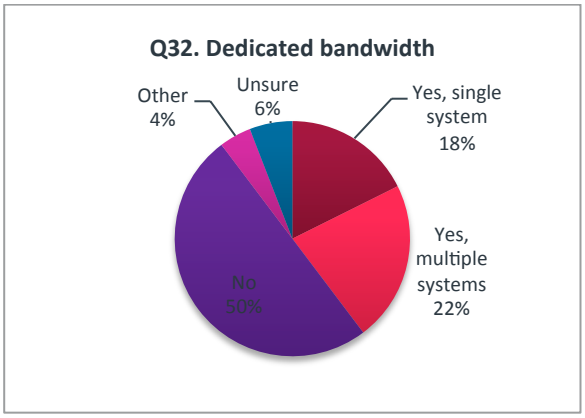
Q30 Peak inbound sustained data transfer rates



Q31 Peak outbound sustained data transfer rates. In most cases this will be the connection speed of the organization overall, but if bandwidth is specifically dedicated for HPC please indicate below



Q32 Is the bandwidth above dedicated for HPC service use?



Q33 Special connectivity requirements e.g. Lightpath, dedicated circuit, low latency

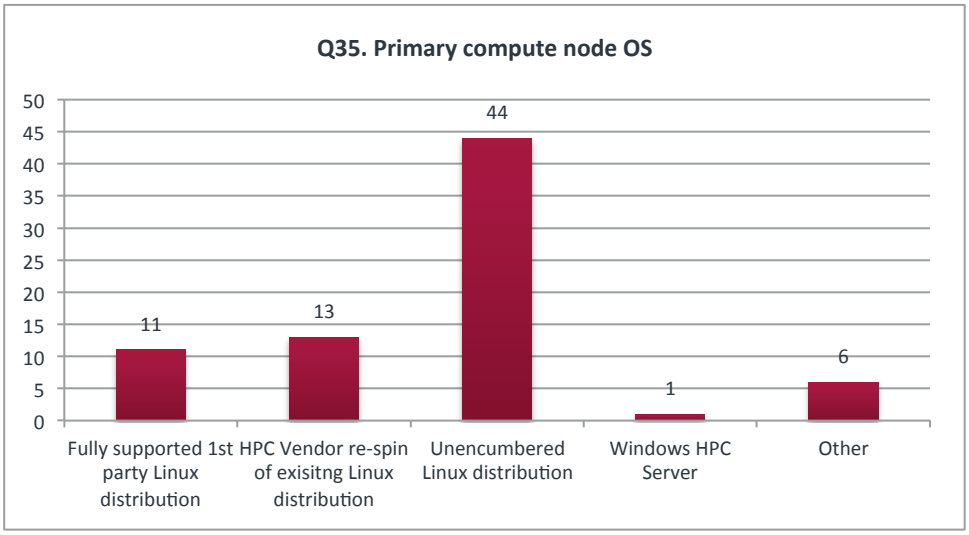
- Replicated storage between sites may in the future require dedicated connectivity. We currently are provisioning ipsec tunnels between sites for secure replication of data.
- Can configure lightpath or dedicated connections as required.
- 10Gb external link to Janet, 40GB internal backplane.
- OPN to CERN at 10Gbits/s. shared 40Gbits/s to Janet.
- The NSCCS machine is a large shared memory machine with 512 Cores and 4TB of memory. This architecture suits the computational chemistry applications.
- OPNs to Met Office Exeter, ARCHER, Leeds University, Space Applications Catapult.
- Lightpath and dedicated networks can be configured on request.
- A function of Application Sector - direct connectivity to NGS systems demands peak performance.
- None at present, but NGS access dictates an order of magnitude increase in connectivity performance.

Software and operating environment

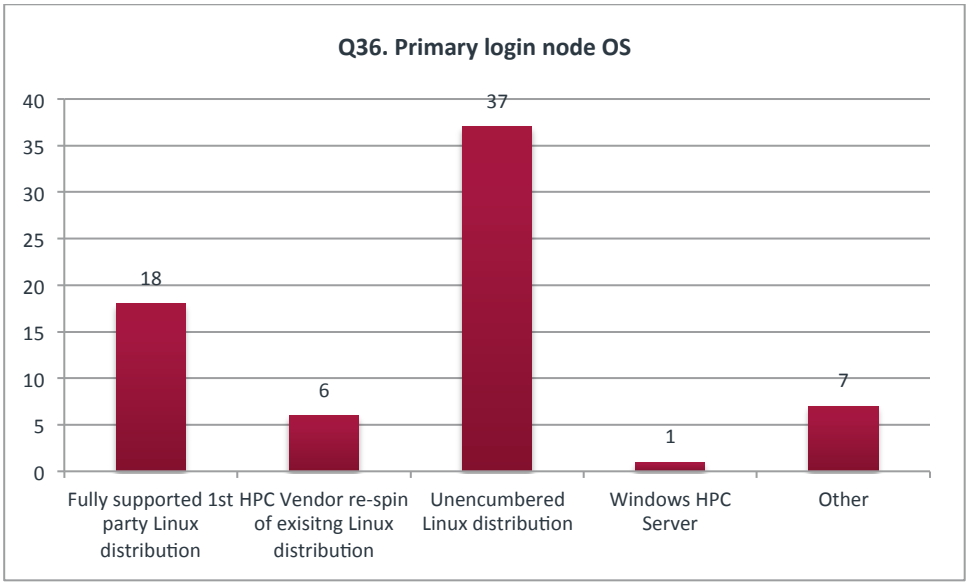
Q34 Installed software

The results from this question were inconclusive due to a bug in the survey tool.

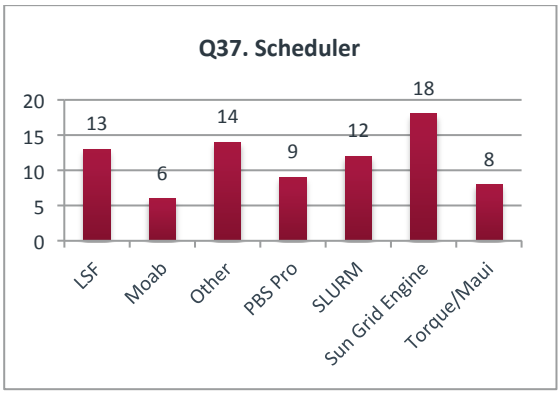
Q35 What is the primary Operating System you use on compute nodes?



Q36 What is the primary Operating System you use on head/login nodes?



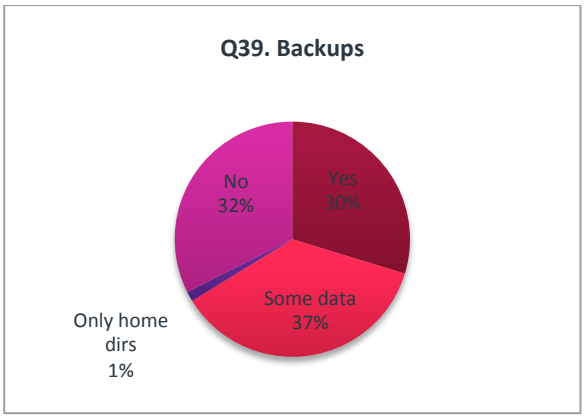
Q37 What scheduler(s) do you use?



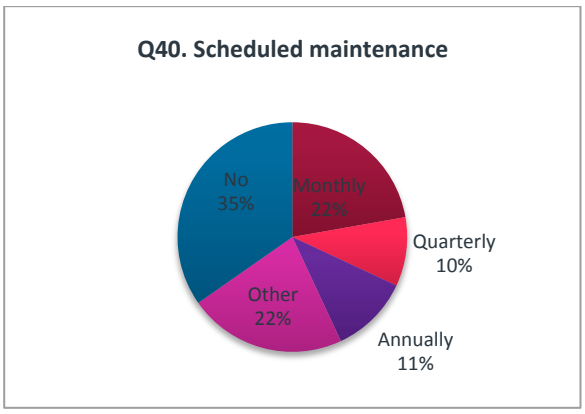
Q38 Do you provide a Web Portal to your users? If so, please describe below

- Horizon is currently only way for most users to access. API access provided via controllers.
- On request
- Yes, SAFE: **https://www.archer.ac.uk/safe** Account Management Project Management Reporting Helpdesk Incident Management System Configuration
- **http://www.cosmos.damtp.cam.ac.uk/**
- Yes - Open Stack Horizon
- AVF is a physical facility where users come and explore their data and interact in 3D. AVF can also offer remote visualisation as well.
- Certificate Wizard Application **http://www.ngs.ac.uk/ukca/certificates/certwizard**
- Yes, needs authentication to use **https://goc.egi.eu/portal/**
- SCARF is accessible via the Platform Application Portal . This allows users to upload and download files, submit computational jobs and design workflows.
- Web access via the Platform Application Portal
- User provisioned cloud web portal available.
- The web portal to the accounting data is provided by CESGA **http://accounting.egi.eu/egi.php**
- Yes, based on Fujitsu's SynfiniWay middleware; see **https://portal.hpcwales.co.uk**
- Yes, for access details and usage information. Not for job submission.
- **http://www.cfi.ses.ac.uk/cfi/iridis/**
- Yes - Galaxy (Life Sciences), WebMO (Chemistry), GridChem (Chemistry) and we are in the process of evaluating both Altair's Compute Manager and Open-Source Cylc (as a multi-purpose web-based interface).
- For gaussian only
- Evaluating Altair Compute Manager, not in production use
- Oracle secure global desktop **http://www.oracle.com/us/technologies/virtualization/secure-global-desktop/overview/index.html**
- **https://maxwell.abdn.ac.uk**- provided by Alces
- Standard BrightCluster Manager with modifications by ClusterVision for Visualisation service

Q39 Do you back up HPC user data?

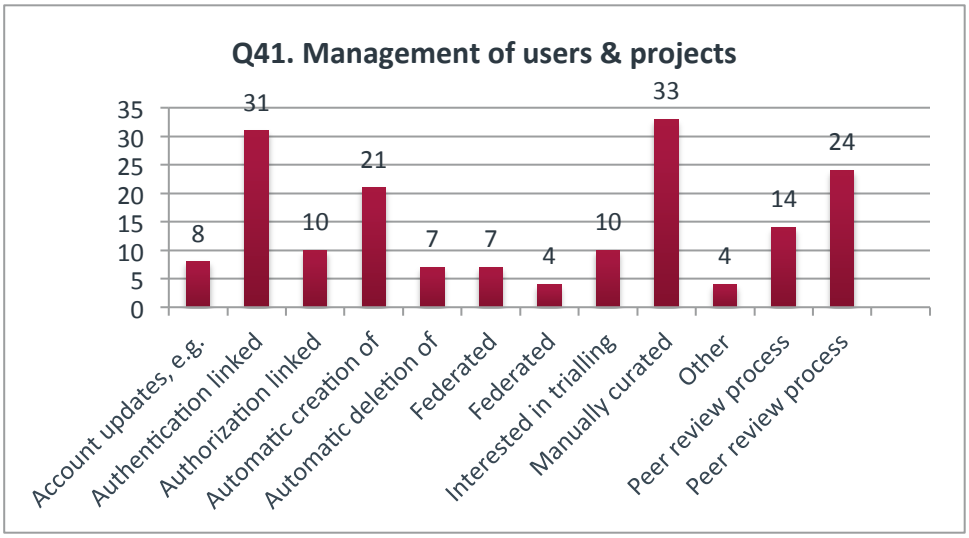


Q40 Do you do have scheduled maintenance and if so how often?

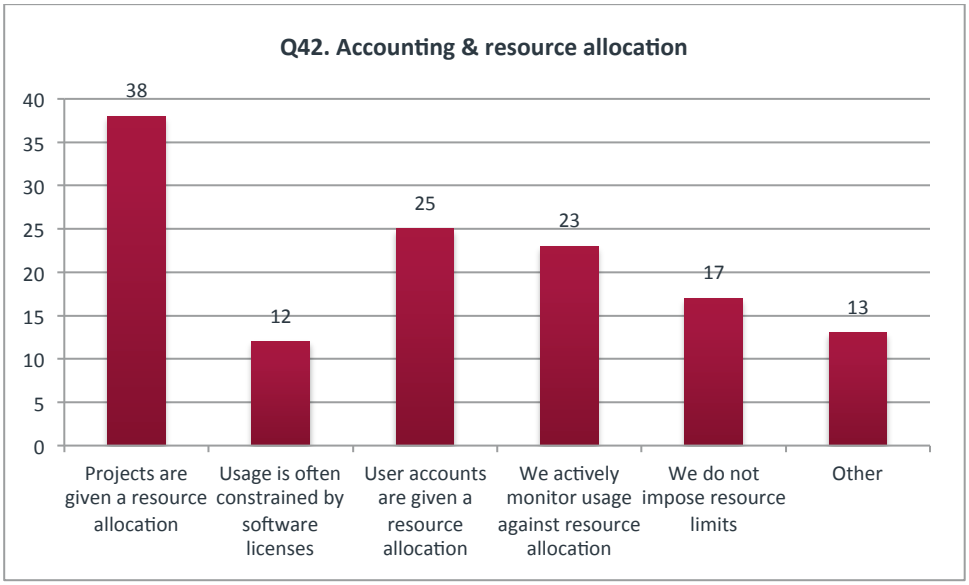


Access, authorization, accounting and identities

Q41 Management of HPC system users and projects



Q42 Accounting and resource allocation



Further information

Q43 Please use this space if there is any other information you would like share about the system

- CLIMB is a different approach compared to traditional HPC systems, so many of the questions don't quite fit! Storage ceph nodes also run RHEL 7.1
- The Computing Infrastructure for Science (CiS) group in NBI Parnership Ltd manages the HPC and enterprise storage for four Institutes on a shared campus network in Norwich: TGAC, JIC, IFR and TSL. The largest proportion of HPC and storage utilisation is from TGAC.
- The AVF is a physical facility with high performance 3D visualisation capability attached to high memory machines for data analysis. The room is used collaboratively with researchers exploring their data together and gaining new insights. The high memory nodes are critical such that large datasets can be manipulated in real time. Remote visualisation is also offered to researchers
- Performance is measured by the HEPSPEC rather than in Tflops as this is more relevant to the workloads. **http://w3.hepix.org/benchmarks/doku.php?id=homepage** The performance of the Tier1 is ~120k HEPSPEC06
- The Atlas Datastore is a hierarchical storage system providing archive, backups , data curation and working repositories.
- The UK e-Science Certificate Authority provides a security infrastructure for people and systems which is accepted globally through the IGTF. The IGTF is the interoperable global trust federation. **http://www.igtf.net**
- GOCDB is the official repository for storing and presenting EGI topology and resources information. The GOCDB data consists mainly of: Participating National Grid Initiatives (NGI) Grid Sites providing resources to the infrastructure Resources and services, including maintenance plans for these resources Participating people, and their roles within EGI operations

- SCARF is managed as part of a portfolio of services that SCD offers to the UK and international science communities.
- NSCCS is part of an collection of services that SCD provides to the UK academic community Over the last year NSCCS has been used by 54 research groups from 24 institutions resulting in over 60 publications.
- Data-intensive computing JASMIN provides the UK and European NERC funded environmental science communities with an efficient data analysis environment. Many datasets, particularly model data, are too big to be easily shipped around: JASMIN enables scientists to bring their processing to the data. Flexible data access JASMIN provides new ways for scientists to collaborate in self-managing group workspaces, enabling models and algorithms to be evaluated alongside curated archive data, and for data to be shared and evaluated before being deposited in the permanent archive.
- APEL is an accounting tool that collects accounting data from sites participating in the EGI and WLCG infrastructures as well as from sites belonging to other Grid organisations that are collaborating with EGI, including OSG, NorduGrid and INFN. The accounting information is gathered from different sensors into a central accounting database where it is processed to generate statistical summaries that are available through the EGI/WLCG Accounting Portal. APEL collects data from ~300 institutions at ~3M records per day totalling ~400GB of data. This is kept for 18 months after which summary data is held. Statistics are available for view in different detail by Users, VO Managers, Site Administrators and anonymous users according to well defined access rights.
- The accounting information is gathered from different sensors into a central accounting database where it is processed to generate statistical summaries that are available through the EGI/WLCG Accounting Portal. The APEL system receives ~3M records per day into a MySQL database from 314 sites. Individual Job records are kept for 18 months (~400GB) . After 18 months the job records are summarised and kept indefinitely. Statistics are available for view in different detail by Users, VO Managers, Site Administrators and anonymous users according to well defined access rights.
- Backup is for disaster recovery purposes only
- A comment to point out the difficulty in entering data for distributed systems - particularly tough in our case e.g., Q19 - could not work out how to enter multiple dates for a multi-phase implementation - 2011 and 2013. One thing to red line a non-acceptable response, but no indication of what constitutes a valid response. Would have been useful to have had a save and restart capability for a questionnaire of this size. Curtailing the input on some of the lines should be accompanied by a maximum filed length.
- Although listed as a single system, there are three partitions - the primary Sandy Bridge partition is the main parallel MPI service, the Westmere is for serial or batch jobs, and the new Haswell service will have a mixed workload (primarily serial but will also be available for MPI-based jobs). This latter partition was installed in January 2015 and the maintenance will expire in Jan 2018 (but unfortunately the maintenance question only permitted a single date entry!).
- Heterogeneous system, so multiple hardware types and support contract spanning 2010 to date. Survey doesn't really accommodate this type of cluster. The cluster is both our local HPC facility and HEP/GridPP node. Funding is a hybrid model - a central University contribution and research grant-funded contributions
- Researchers have to register a project, which may be unfunded, and users have to be associated with a project. This enables linking research students to their supervisor which is very useful in case of problems. It also enables us to identify the range of users and disciplines which is essential to making the case to the University for ongoing funding.
- We are about to undergo a software upgrade that will rather change the OS and application availability (for example we are moving to RHEL from SL). Commission date is inaccurate as this system has been in use for a long time. It is also not possible to give a maintenance cutoff because that varies from already expired to brand new equipment.

- I skipped the questions on commissioning and maintenance dates, because we upgrade on an annual cycle, and once the original support agreement expires we often roll the HPC node into our campus maintenance agreement with HP.
- We also have an active Condor system with circa 20 active users, running across 750 campus PCs, that typically accommodates 20,000 core days of computing per month via approximately 60,000 jobs (per month). This relieves pressure off the HPC system for high volume, high throughout jobs.
- The system is heterogeneous and based on a contribution model. Central seed funding bought infrastructure. Research groups contribute funds to buy compute nodes. Procurements take place two or three times a year.
- The University funds a general HPC service which is augmented by the addition of research grant or school funded compute and storage. Nodes and Storage are purchased with 5 years warranty, after which they are considered End of Life. We use Bright for Cluster provisioning and management.

Appendix J – Hardware: Summary of the survey data

Table: The hardware systems are given in the Table below which is split to into 3 sections, each representing a layer in the Branscomb Pyramid. Section 1 lists the Large and Specialist Systems. Section 2 gives the Regional Systems. Section 3 lists the foundational layer, the HEI sector.

1. Large and Specialist Services						
Organization name	**System name**	**What are the top three research areas the system is used for?**	**Total number of processor cores in the system**	**Total usable storage for HPC users (TB)**	**Number of registered users**	**Theoretical Peak Performance (Tflop/s)**
Cloud Infrastructure for Microbial Bioinformatics - CLIMB - Birmingham, Cardiff, Swansea, Cardiff	CLIMB	Research in how microbial bioinformaticians use cloud Microbial bioinformatics (assembly, analysis) Building an academic cloud	3,864		0 - 100	74
DiRAC @ Durham University	DiRAC-1@Virgo (COSMA4)	Cosmology Galaxy Formation	7,072	1,100	100 - 200	32
DiRAC @ Durham University	DiRAC-2@DataCentric (COSMA5)	Cosmology MHD Galaxy Formation	6,720	2,500	100 - 200	140

DiRAC @ EPCC	DIRAC BG/Q	QCD, Soft Matter Physics	98,304	1,000	200 - 500	1,258
DiRAC @ University of Cambridge (DAMTP)	COSMOS	Cosmology, Astrophysics, Exoplanets	3,284	306	200 - 500	63
DiRAC @ University of Cambridge (HPCS)	Darwin	Life Sciences. Atomic structure. Computational Fluid Dynamics.	9,600	2,847	750 - 1,000	200
DiRAC @ University of Leicester	Complexity	Astrophysics Particle physics	4,352	710	100 - 200	91
EMBL-EBI - European Bioinformatics Institute	Embassy Cloud	Life science research	31,000	3,200	200 - 500	
eMedLab - The Crick, UCL, QMUL, LSHTM, EBI, The Sanger	eMedLab	Biomedical research (and any associated areas). Next generation sequencing RNA sequencing Cryo-electron microscopy	6,048	4,800	0 - 100	
UVRI/MRC Medical Informatics Centre	UMIC	Medical Informatics (not yet in operation)	2,048	1,720	0 - 100	19
Norwich Bioscience Institutes (TGAC, JIC, IFR, TSL)		Bioinformatics, mathematical modelling.	9,000	4,000	750 - 1,000	
STFC Hartree Centre	Blue Joule	Modelling & Simulation (CFD, Materials, and Computer Aided Formulation)	98,000	6000	200 - 500	1,200

STFC Hartree Centre	BGAS	Modelling & Simulation (CFD and Materials)	32,000	1000	0 - 100	450
STFC Hartree Centre	Blue Wonder	Modelling & Simulation (CFD, Materials, and Computer Aided Formulation)	24,000	9000	750 - 1,000	200
STFC Hartree Centre	EECR/FPGA	Energy Efficient Computing	2,000	200	0 - 100	
STFC Hartree Centre	BigData	Big Data and Data Analytics	1,184	1000	0 - 100	
STFC Scientific Computing Division	SCARF	Computational Chemistry Plasma Physics, Processing Satellite images Support of ISIS, CLF, RAPSP, DLS user communities	7,000	320	500 - 750	165
STFC Scientific Computing Division	JASMIN	Climate Science, Earth Observation, environmental genomics	4,500	25	Over 10,000	

STFC Scientific Computing Division	National Service for Computational Chemistry Software	The EPSRC UK National Service for Computational Chemistry Software (NSCCS) provides access to software, specialist consultation, computing resources and software training to support UK academics working across all fields of chemistry. NSCCS supports 127 research groups from disciplines including Chemistry, Materials Sciences, Physics, Earth Science and Engineering, Astronomy, Biochemistry, Biological Sciences, Biomedical Sciences, Life Sciences, Civil and Environmental Engineering and Chemical Engineering	512	32	200 - 500	
STFC Scientific Computing Division	Atlas Visualisation Facility (AVF)	Material science tomography and Plasma Physics	112		0 - 100	
STFC Scientific Computing Division	Atlas Datastore	Particle Physics, Life Sciences, Materials Science plus everything else		25,000	100 - 200	
STFC Scientific Computing Division	UK e-Science Certfication Authority	Supports all UK research. Major users Particle Physics			750 - 1,000	
STFC Scientific Computing Division	GOCDB	Supports multi-national e-infrastructures			2,000 - 5,000	

STFC Scientific Computing Division	GridPP Tier1	Particle Physics Expanding to support other disciplines	10,000	28,000	1,000 - 2,000	
STFC Scientific Computing Division	APEL	APEL is a central repository for accounting/usage data. It supports local, grid and cloud computing.				
STFC Scientific Computing Division	APEL	APEL is an accounting tool that collects accounting data from sites participating in the EGI and WLCG infrastructures as well as from sites belonging to other Grid organisations that are collaborating with EGI, including OSG, NorduGrid and INFN. APEL currently supports local, grid and cloud data			200 - 500	
The Institute of Cancer Research	Multiple	Processing of sequencing, mass spec and imaging data.	1,600	2,500	0 - 100	
EPCC	ARCHER	Materials Science, Climate/Ocean Modelling, Computational Fluid Dynamics	118,080	4,608	1,000 - 2,000	2,550
EPCC	Indy	Industry/SME Use	1,536	175	100 - 200	
EPCC	Ultra	Healthcare, Bioinformatics	512	253	0 - 100	

EPCC	DIR	Data analytics	240	2,250	0 - 100	
EPCC	UK-RDF	Climate/ocean modelling, Computational Fluid Dynamics, Materials Science		23,000	1,000 - 2,000	
Farr North, Health eResearch Centre	HeRC Safe Haven	Healthcare Bio-health Informatics Machine Learning	256		0 - 100	
Wellcome Trust Sanger Institute	Sanger HPC Resources	Genomics	16,946	9,900	1,000 - 2,000	
LARGE & SPECIALIST TOTALS			499,770	135,446	30,550	6,441

2. Regional Systems						
Organization name	**System name**	**What are the top three research areas the system is used for?**	**Total number of processor cores in the system**	**Total usable storage for HPC users (TB)**	**Number of registered users**	**Theoretical Peak Performance (Tflop/s)**
High Performance Computing (HPC) Wales	Various (distributed system)	Advanced Materials & Manufacturing, Life Sciences and Energy & Environment	16,816	702	2,000 - 5,000	319
HPC Midlands	Hera	Advanced Materials Energy Efficient Transport	3,008	120	100 - 200	48
N8HPC	Polaris		5,312	175	200 - 500	138

ARCHIE-WeSt	ARCHIE	Molecular dynamics, CFD, Plasma Physics	3,920	148	200 - 500	38
SES/CFI	IRIDIS3	Chemistry research, Engineering research and Biology research	12,000	110	200 - 500	106
REGIONAL TOTALS			41,056	1,255	49,660	649

3. HEI Systems

Organization name	System name	What are the top three research areas the system is used for?	Total number of processor cores in the system	Total usable storage for HPC users (TB)	Number of registered users	Theoretical Peak Performance (Tflop/s)
Cardiff University	Raven	EPSRC (materials, chemistry, engineering), BBSRC (genomics) and NERC (earth sciences)	4,352	275	500 - 750	110
Cranfield University	Astral	CFD FEA	1,280	34	200 - 500	20
Durham University	Hamilton	Condensed Matter Molecular Dynamics Fluid Dynamics	5,600	350	200 - 500	75
Imperial College London	ax3	Genomics	1,300	1,500	0 - 100	
Imperial College London	cx1		21,558	2,000	750 - 1,000	
Imperial College London	cx2		7,000	500	0 - 100	60

King's College London	ADA	Computational Physics, Mathematics, Informatics	1,624	87	100 - 200	85
Lancaster University	HEC (High End Cluster)	High Energy Physics Condensed Matter Theory CFD	4,784	1,530	200 - 500	
Loughborough University	Hydra		2,460	100	100 - 200	
Queens University of Belfast	Dell Cluster	Chemistry, Physics, Cancer Research	984	40	0 - 100	
Queens University of Belfast	Windows Cluster	biology, speech and image processing, CFD	256	30	0 - 100	
The University of Birmingham	bluebear.bham.ac.uk	Mathematics Civil Engineering Pyschology	1,632	150	500 - 750	21
The University of Nottingham	Minerva	Chemistry, Engineering, Physics	2,752	180	200 - 500	55
The University of Sheffield	iceberg1	Aeronautical Engineering and modelling of turbulent fluids Computational Medicine Bioinformatics	3,440	40	1,000 - 2,000	112
University College London	Legion	Chemistry, Physics, Biological Sciences (according to REF Categories)	7,816	356	500 - 750	115
University of Aberdeen	maxwell.abdn.ac.uk	Life Sciences - Genomics CFD - Engineering Matlab	600	56	100 - 200	

University of Bath	Balena / Aquila	Chemistry Physics Mechanical Engineering	3,072	220	0 - 100	64
University of Bristol	BlueCrystal	Chemistry, Aerospace Eng, Geographical Sciences	9,000	740	750 - 1,000	240
University of Cambridge	Wilkes	Computational Fluid Dynamics. Atomic structure.	1,536		0 - 100	256
University of Edinburgh	Eddie	Physics, Informatics, Engineering	3,248	281	1,000 - 2,000	28
University of Exeter		Astrophysics Weather/climate Hydrology	2,184	73	100 - 200	25
University of Glasgow	Conan	Semiconductor Device Modelling	1,360	40	0 - 100	
University of Glasgow	Cnoc	Electronic System Design	320	10	0 - 100	
University of Glasgow	Dusty	Computational Fluid Dynamics	188	4	0 - 100	
University of Glasgow	Miffy	Semiconductor Device Modelling Computational Fluid Dynamics Optoelectronics	1,256	22	0 - 100	15
University of Leeds	Arc1	CFD, Astrophysics, climate science	4,128	117	500 - 750	31
University of Leeds	Arc2		3,040	175	200 - 500	316

University of Leicester	ALICE	Astrophysics (now largely migrated to DiRAC) Earth Observation Science Engineering Economics	3,972	1,644	200 - 500	101
University of Liverpool	chadwick	Materials modelling Computational Fluid Dynamics Gait Analysis	3,180	132	100 - 200	23
University of Manchester	Computational Shared Facility	Computational Chemistry / MD CFD FEA	6,288	750	750 - 1,000	111
University of Oxford	Arcus-A		1,728		2,000 - 5,000	55
University of Oxford	Arcus-B		5,440	432	2,000 - 5,000	538
University of Oxford	Arcus-GPU		12		2,000 - 5,000	146
University of Portsmouth	Sciama	fundamental physics, cosmology and astrophysics	3,704	740	100 - 200	
University of St Andrews	wardlaw	MHD, Astronomy, Chemistry	3,510	150	200 - 500	33
University of Sussex	Apollo	Physics (Astronomy, Cosmology,Particle), Engineering (CFD), Informatics (Computational Neuroscience, Adaptive systems, Natural Language)	3,248	560	100 - 200	
HEI TOTALS			127,852	13,318	30,430	2,635

	Total number of processor cores in the system	Total usable storage for HPC users (TB)	Number of registered users	Theoretical Peak Performance (Tflop/s)
GRAND TOTALS	1,209,504	286,720	67,680 (upper bound)	16,815