Science & Technology Facilities Council

Computing Strategic Review

December 2015

Front cover – bottom right image shows a view of dark matter at redshift zero in the Eagle simulation, by the Virgo Consortium using the DiRAC Facility (Durham Data Centric Cluster)

STFC Computing Strategic Review

Contents

Foreword 1

Executive Summary 2

Introduction 4

Vision and Objectives 7

STFC Programme Areas: Current Usage, Future Plans, Issues and Opportunities 9

- Particle Physics, Astronomy and Nuclear Physics Research 9
- The UK National Facilities 14
- Engineering Support 17
- Externally Funded Projects and Services 20
- Hartree Centre 24

Cross-Programme Topics 27

Recommendations 33

Annex 1: The Strategic Review Development Process 37

Annex 2: Glossary 39

Annex 3: Summary of Responses to the Survey of current computing facilities and services 40

- Part 1 - Facilities, Scientific Computing Department and the Hartree Centre 40
- Part 2 – Particle Physics, Astronomy and Nuclear Physics Research Programme 50

Foreword

The Science and Technology Facilities Council is one of Europe's largest multidisciplinary research organisations. Computing and data handling underpin everything that we do. The world-leading science we support through research at universities and our large-scale facilities depends on our computing facilities, services and internationally recognised expertise. Experiments need to transfer, analyse and store ever increasing amounts of data and this requires computing hardware, software, networks and skilled professional support.

Computing also plays a central and growing role in our innovation activities. The Hartree Centre uses its software expertise and supercomputing power to deliver a competitive edge for UK business.

In the future it is clear that the pace of technological change will continue and the demands placed on our computing facilities and services will dramatically increase. We need to ensure that we are in the best position possible to meet these challenges so that we can achieve our strategic goals of delivering world-class research, innovation and skills.

This Strategic Review assesses our current computing capability and future requirements across our entire programme. It recommends actions to meet our future computing needs and improve efficiency. The Review provides the framework within which we will develop a strategic plan for computing. Following the publication of this Review, we will consult our communities to evaluate how we can best implement the recommendations within the funding available and decide our future priorities for computing. In the years ahead this should ensure we can provide the optimum computing support to our science and innovation communities.

Professor John Womersley
Chief Executive Officer

Executive Summary

Computing is vital to deliver STFC's strategic goals of delivering world-class research, innovation and skills. The STFC Computing Strategic Review is an internal STFC review that aims to identify the current and future e-infrastructure requirements of STFC's scientific and industrial user communities and recommend how STFC can best support the provision of those needs. The review panel consulted with the users of STFC computing and other stakeholders to develop its recommendations. The Review covers hardware, data services, software, cyber security, networks and computing professionals.

The vision for STFC Computing is

'to increase the scientific, technological or economic output of users of STFC computing by exploiting technology, developing skills and optimising the use of resources'.

The users of STFC computing comprise STFC-funded researchers; users of STFC National Facilities; other science communities supported by computing facilities or services hosted by STFC; and clients of the Hartree Centre.

All STFC programme areas anticipate an order of magnitude increase over the next five years in data volumes, with implications for requirements for computing hardware, storage and network bandwidth. Meeting this challenge is essential to ensure the UK can continue to produce world-leading science, but in the current financial climate it is clear there will be funding limitations.

Long-term planning will be critical, as will more efficient use of resources. The scientific disciplines supported by STFC should work more closely together to find ways of sharing computing infrastructure, with new projects encouraged to make use of existing expertise and infrastructure. STFC should play an active role in co-ordinating this, and maintain good communications with partner organisations such as the UK Space Agency on future project compute requirements. Consolidation is possible, while recognising that users require facilities ranging from desktops to the largest Tier 0 machines, for diverse specialities and at various locations.

STFC needs to facilitate access to its compute resources, including professional expertise, by clearly communicating the resources available and providing appropriate user guidance and support. Interoperability and common authentication, software, metadata and file formats etc. will increase the effectiveness of the resources available.

The growing diversity among emerging computing architectures will also affect all areas. STFC and its expert user communities should make full use of their computing capability to provide leadership in making the best use of developments in compute technology. STFC should also ensure that it maintains the skills and capabilities to support the large-scale scientific computing facilities that are necessary for a broad UK science base.

Data scientists and software engineers are becoming vital to all research areas and need to be supported by ensuring there is a clear career path for these professionals both within STFC and at the universities.

Computing provision for National Facility users should be reviewed as an urgent priority, including high performance computing, software, expertise, data storage, data transfer to facility user sites, and access to archived data at the facilities (providing simple remote access for non-expert users).

The Hartree Centre is set to play a crucial role in boosting UK business, with a government partnership with IBM providing a £315M investment in its data centric and cognitive computing capabilities. As plans for this exciting development progress over the coming months, the Hartree Centre's position in STFC's governance structure and corporate strategy, its role and remit, needs to be communicated clearly to all stakeholders.

Introduction

Scope of the Strategic Review

Computing is a cross-cutting activity that supports all STFC's strategic goals of delivering world-class research, innovation and skills. It is vital to all research supported by the Science and Technology Facilities Council (STFC). This includes the particle physics, nuclear physics and astronomy research programmes (both in the universities, our laboratories and at the international facilities); the experiments carried out by users of the STFC National Facilities; and the scientific projects funded by other Research Councils who rely on computing facilities and services hosted and operated by STFC. Computing also plays a key role in delivering STFC innovation, most notably at the Hartree Centre where government investment in high performance computing systems is intended to give UK business a key advantage over competitors.

The scope of the Computing Strategic Review covers all of the above. STFC computing resources to support its own administrative needs are out of scope of this review.

The Review addresses e-infrastructure in its widest sense: hardware, data services, software, cyber security, networks and computing professionals.

Aims and purpose of the Strategic Review

The Strategic Review aims to:

- identify current and future e-infrastructure requirements of STFC's scientific and industrial user communities
- recommend how STFC can best support the provision of these needs
- identify the relation to the wider UK and international setting
- consider how to maximise the impact of STFC e-infrastructure both for science and for industrial engagement.

The Strategic Review will be used:

- to inform STFC's decision-making on e-infrastructure
- to inform STFC science communities, other Research Councils and stakeholders of STFC's priorities for e-infrastructure
- as a resource for preparing for spending reviews and responding to funding opportunities
- to inform the Scientific Computing Department's (SCD) strategy for the development of new compute infrastructure and service.

The STFC Computing Strategic Review in context

Computing is a key enabler of STFC science. The highest priority current and future scientific projects identified in the 2013 STFC Programmatic Review, such as the Large Hadron Collider (LHC) experiments, the European Extremely Large Telescope and gravitational wave experiments, will depend on computing to collect, process and store data. Access to state-of-the-art computing facilities will be vital to ensure UK science remains internationally competitive.

In recent years e-infrastructure has gained prominence on the government's science and innovation agenda: Big Data is one of the Eight Great Technologies; the E-Infrastructure Leadership Council was created in 2012; and the UK Data Capability Strategy was published in 2013. Big Data was identified as a key priority in the 2014 *Government Response to Consultation on Proposals for Long-term Capital Investment in Science and Research*. Recent years have seen substantial government capital investments in e-infrastructure projects, including most recently £115m in 2014 to expand cognitive computing capability at the Hartree Centre. This investment is leveraging access to a further £200m of intellectual property from IBM, giving a total value of £315m

STFC computing is provided in concert with a range of organisational partners: funders such as government, other Research Councils, the UK Space Agency; commercial partners and collaborators; infrastructure support e.g. Jisc for the Janet network, universities; collaborators in research projects etc. The international nature of many projects means the computing also has an international dimension: project collaborations can span many countries and depend on international bodies such as the European particle physics laboratory (CERN) and the European Southern Observatory (ESO). For example, the UK particle physics computing grid, GridPP, is part of the Worldwide LHC Computing Grid, a global collaboration of 41 countries. STFC scientists use international facilities such as the Partnership for Advanced Computing in Europe (PRACE) supercomputing facility.

Technology continues to evolve at a rapid pace, with the development of new computer architectures and models such as massively multi-core technologies and cloud computing. Future challenges include increased power costs as system performance increases, driving the need for more energy-efficient computing. As mathematical and scientific models grow in complexity they require much more complex codes. The massive increase anticipated in data volumes presents a significant challenge: the Square Kilometre Array (SKA) radio telescope predicts it will need to process 10,000 petabytes per day by 2028. However, developing the novel technology to meet such challenges also provides innovation opportunities: solutions may have applications for other research projects and there is potential for engagement with industry. The exploitation of large diverse data sets can also create knowledge and wealth, with the Hartree Centre's expertise and capability in big data analytics providing a competitive edge to a variety of industrial clients. The Review will address these opportunities, risks and challenges.

STFC has a uniquely diverse remit, supporting research programmes in particle physics, nuclear physics, astronomy and exploitation of the space science missions funded by UKSA; providing National Facilities (Diamond Light Source (Diamond), Central Laser Facility (CLF), ISIS (neutron and muon source)) to support a broad spectrum of scientific communities; and developing the UK Science and Innovation Campuses at Harwell and Daresbury. The provision of computing support across the STFC programme is therefore complex, with some computing infrastructure being provided directly by STFC, and other infrastructure being funded by STFC but delivered by the universities and/or international organisations of which we are members. The Scientific Computing Department not only provides essential computing support for the National Facilities but also utilises its technical expertise to support projects and services of other Research Councils. This Review attempts to provide a cross-STFC overview of future computing needs and opportunities to increase efficiency and effectiveness of our computing provision.

The Computing Strategic Review development process

The Strategic Review was carried out by an internal review panel with representation from the Computing Advisory Panel. The review process included a survey of STFC's current computing facilities and services and a consultation with key stakeholders to identify future requirements and opportunities for increased efficiencies and effectiveness. Documents such as strategies and policies from relevant organisations were reviewed to ensure alignment with national and international priorities. Annex 1 describes the review development process in more detail.

Following publication of this review, STFC will consult our communities to consider how the review's recommendations should be implemented. This will require evaluating the implications of implementation and deciding priorities where funding is required.

Vision and Objectives

The vision and objectives for STFC computing over the next five years (2015 – 2020) are set out below.

The Strategic Review is divided into two main sections: one covering the principal programme areas covered by the Review and one which describes topics and issues which cut across multiple programme areas. These sections contain detailed recommendations for actions that will contribute to the delivery of the vision and objectives.

Vision

To increase the scientific, technological or economic output of users of STFC computing by exploiting technology, developing skills and optimising the use of resources.

The users of STFC computing comprise:

- STFC-funded researchers
- Users of STFC Facilities (CLF, Diamond, ISIS)
- Other science communities supported by computing facilities or services hosted by STFC
- Clients of the Hartree Centre.

STFC computing covers e-infrastructure in the broad sense of hardware, data services, software, cyber security, networks and computing professionals.

Objectives

1) Support the STFC strategic goals of delivering world-class research, innovation and skills
2) Develop a balanced approach to supporting computing requirements that vary from small-scale to large-scale, generic to specialised, local to centralised
    a) Provide clear, accessible information on the resources available and guide users to the most appropriate resource
    b) Equip users with appropriate tools, services, support or training to make the best use of STFC resources
3) Develop a strategic and co-ordinated approach to planning for future computing requirements
    a) Ensure the development of affordable planning provision for computing infrastructure requirements
    b) Ensure that planning for the computing requirements of projects starts at the earliest possible stage of project planning
    c) Put in place mechanisms to identify opportunities where sharing resources would be of benefit (between projects, science disciplines, funding agencies etc.)
    d) Provide an overview so that the impact of proposed changes can be assessed across the organisation
    e) Ensure that STFC's role as a provider of resources to scientific communities outside STFC-funded programmes and Facilities is appropriate in terms of meeting the needs of the communities and making the best use of STFC resources

4) Develop and make the best use of the computing expertise within STFC or the communities it supports
    a) Put in place mechanisms to facilitate the sharing of expertise and best practice
    b) Put in place measures to support the development of a career path for software engineers/data scientists in a research environment
    c) Provide guidance on the adoption of cutting-edge and future technologies
5) Improve support for data management, analysis, transfer, storage and publications, specifically:
    a) Promote the appropriate use of open data
    b) Prepare for increasingly challenging data requirements by ensuring optimal use of resources (code optimisation; investment in software; fast, robust networks etc.)
    c) Assist users to make use of the most appropriate software

STFC Programme Areas: Current Usage, Future Plans, Issues and Opportunities

Particle Physics, Astronomy and Nuclear Physics Research

a) Current usage

STFC provides e-Infrastructure resources across the particle physics, astronomy and nuclear physics (PPAN) programme. This includes computing hardware, storage, networks, and the support of computing personnel.

STFC provides computing resources over a range of scales to service the needs of the community. The largest STFC resources are the Tier-1 particle physics compute centre, funded through GridPP (the UK particle physics computing grid), and hosted by the Rutherford Appleton Laboratory (RAL), and the various DiRAC (Distributed Research utilising Advanced Computing) facility sites providing high performance computing. These provide computing resources of tens of thousands of cores, tens of petabytes of data and/or tens of terabytes of shared memory. The next level is comprised of smaller, Tier-2 sites at universities, such as those funded through GridPP, or astronomy grants, which are typically an order of magnitude smaller. The smallest scale, Tier-3 computing, comprises desktop computing which is critical for interpretation and visualisation of results obtained at the Tier-1 and 2 centres. Although STFC does not provide any Tier-0 resource, other than through the CERN subscription, STFC funded researchers have been successful in securing time on the European PRACE infrastructure, which offers machines with hundreds of thousands of cores. Finally, several projects, such as Galaxy Zoo and LHC@home, have used citizen science and volunteer donated computing to fulfil their computing requirements.

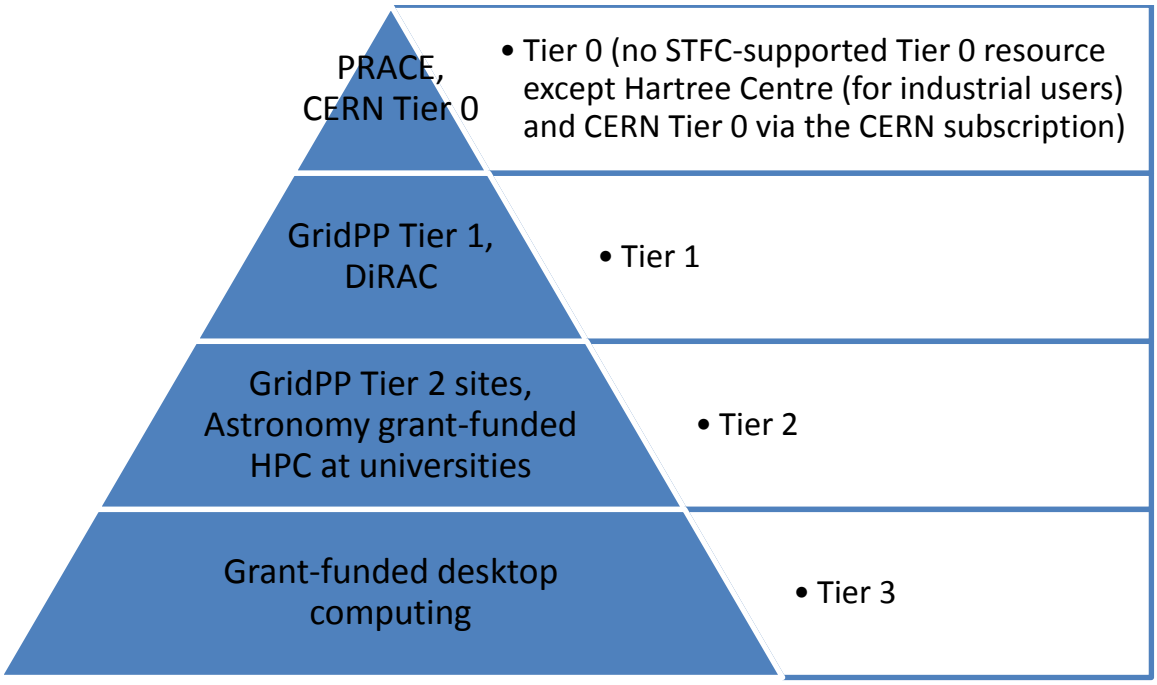

Figure 1: The different scales of computing resources.

These resources currently provide the appropriate computing for the PPAN community. In cases where there is a concern about lack of resources, it is often due to a lack of readily available information: descriptions of available resources, their current usage and procedures to gain access. While available HPC facilities are adequate, they are significantly smaller than what is available in other countries, notably the United States. Nonetheless, by using these resources effectively, the UK has generated world-leading scientific output in numerous areas, including experimental and theoretical particle physics as well as astrophysics simulations of stars and galaxies. The majority of active users rely on access to Tier 1, 2 and 3 computing with large scale runs performed on Tier 1 centres, development and test runs performed on Tier 2 sites and data visualisation and interpretation performed on Tier 3 machines. The continued availability of all levels of resource, at an internationally competitive level, is required to maintain the UK's scientific excellence.

Network connectivity between sites in the UK is provided by Jisc. This encompasses the national network (Janet) augmented by a handful of dedicated links, such as those from CERN to the Tier 1. International connectivity is routed through Geant.

No current gaps in provision have been identified, although all areas of PPAN science highlighted the expected growth in e-Infrastructure requirements over the coming years.

b) Currents plans; long term requirements

Most areas of PPAN research expect their computing needs to grow over the coming years as data sets get larger and more complex. The general view is that at least one of: processing requirements, size of data sets or network bandwidth will grow by an order of magnitude for existing projects. In addition, there are several new instruments in astronomy that will drive this rapid expansion, most notably the SKA, the Large Synoptic Survey Telescope (LSST) and the Cherenkov Telescope Array. These will begin to challenge the existing networks as well as the data and compute requirements.

There are a small number of users who require Tier-0 scale computing. At present, this is not readily available to STFC researchers other than through applications to PRACE. The UK does have a Tier-0 scale machine, ARCHER, which is managed by the Engineering and Physical Sciences Research Council (EPSRC) and improved access for STFC researchers would help this group.

There are plans in place to expand the Tier-1 level computing with proposed expansions to GridPP (GridPP-5) and DiRAC (DiRAC 3) which, if they are approved and fully funded, would go a long way towards meeting the increased computing requirements over the next three or four years. In the absence of government investment, DiRAC 3 may not be feasible. A number of the new projects will require computing that is close in scale to the existing Tier-1 centres. In many cases, this could be provided more effectively by an expansion of the existing centres rather than establishing new project specific centres.

There is a continued need for the Tier-2 level machines where the smaller scale analyses are run. These typically provide a faster turnaround and more flexible service than can be achieved at the national facilities. The universities that host the Tier-2 machines often provide a significant amount of additional support and future computing models need to recognise the advantages of this additional leverage. Finally, there will continue to be the need for desktop scale computing. Even

within the High Throughput Computing (HTC) and High Performance Computing (HPC) communities, the final stage of data visualisation is typically performed on a local machine. Support for visualisation, as well as fast transfer of data products, continues to be a priority.

The expected ten-fold increase in data will require an increase in network capacity. This will likely require a number of new, dedicated links (such as for SKA and LSST and also to/from the Tier-1 centres) as well as an overall improvement of the network. There are not any significant feasibility concerns regarding this expansion, as the evolution of Janet is expected to keep pace with these requirements.

The majority of the science within PPAN is done in an international context and the computing resources that are provided must reflect this. International collaborations typically have developed a complex set of software and middleware that will need to be deployed on the UK provided resource. In addition, they will have requirements for authentication and authorisation for collaboration members.

c) Issues and opportunities

The continued growth of e-Infrastructure requirements across the whole of PPAN will require a coherent approach to the provision of computing hardware, software and personnel. Scientific computing is a long-term project, whether it is in support of an experiment or in producing the next generation of simulations. Thus, stability in funding over the longer time-scale must be available to allow for long-term planning of the computing requirements of a project. Consequently, it is important that computing requirements are considered at the earliest stages of project planning, even if detailed, accurate costings are not available until the planning is at a more advanced stage.

The existing computing tiers remain appropriate for provisioning of resources, and support for all levels of computing remains important. Issues with availability of computing are more often due to a lack of knowledge of existing systems rather than availability of hardware. STFC could assist by better publicising the availability of resources. Additionally, STFC should provide clear guidelines detailing the scale of computing that is appropriate for each tier, as well as better documentation on procedures to access resources. Greater transparency on funding bids and the process of time allocation on Tier-1 machines would also be beneficial.

One of the major challenges facing HPC and HTC users is a change in available computing hardware. The speed of individual cores has reached a plateau and instead the number of cores per machine is rapidly increasing. Furthermore, we are witnessing the rise of specialist hardware such as Graphics Processor Units (GPUs) and Intel MiCs. To make optimal use of this hardware requires significant effort in porting the code to a new platform and undertaking detailed optimisation work. In many cases, this is not an area where the project scientists have expertise. There is an opportunity to co-ordinate this effort at the national scale.

There is a continued pressure towards open data and this leads to significant requirements on data curation and preservation. Currently, data curation and preservation is carried out to some extent (for some data sets) at STFC sites, but it is not something that is handled by the DiRAC centres. A unified strategy to data preservation will be critical in the coming years. A key part of this will be the

generation and publication of the appropriate metadata that will allow researchers and the public to find and access the appropriate data sets.

The provision of computing hardware, storage and networks is only useful if accompanied by expert computational support that allows researchers to make use of this e-infrastructure. In many cases, there are excellent software engineers working to ensure this is the case. However, the career path for software engineers and data scientists is by no means clear, and the long-term prospects for computing support staff need to be improved. Additionally, it is critical to situate these people appropriately. In some cases it is preferable to site computing staff close to researchers to allow for regular interaction. In other instances, a centralised hub is more appropriate with the computing specialists managing the centralised, specialised functions, and talking with the computing specialists embedded in the domain specific "satellite". These two models work very effectively together, and both need supporting.

Meeting the future computing needs of existing and new research projects will present major technological challenges. But overcoming these challenges will offer possibilities for industrial engagement and innovation opportunities. Some of the novel technology developed will almost certainly have applications to other research problems. STFC should facilitate partnerships between academia and industry that will enable both sides to benefit from each other's expertise. There will be business opportunities for UK companies to provide hardware, software and skills.

d) Recommendations

The PPAN research planned in the next five years requires an order of magnitude increase in computing hardware, storage and network bandwidth. STFC should provide a plan to ensure this is made available. This can be achieved by maintaining and strengthening the existing hierarchy of computing (from Tier-1 to Tier-3), whilst ensuring access to a diverse set of compute resources (high memory, data intensive, high throughput, massively parallel). STFC should ensure that access to Tier-0 scale computing remains available to those researchers who require it. The available resources should be better publicised and guidance provided to ensure that projects are making use of the most appropriate resources. When required, it is important to provide assistance to people moving up (or down) the tiers. Furthermore, STFC should take a more active role in co-ordinating use of e-Infrastructure and investigating the opportunity to share resources, without sacrificing the variety and specialist resources required by various communities. This is particularly relevant for new projects, which should be strongly encouraged to make use of existing expertise and infrastructure such as data centres, although it is important to ensure this is a joint venture and projects are not pushed onto inappropriate infrastructure. Computing should be integrated into the long-term planning of projects. This will assist with the development of well-defined career paths in scientific computing.

Improved authentication and authorisation infrastructure, which is standardised across the community, would facilitate the sharing of compute resources and access to data, applications and services. Cloud computing can provide researchers with simple, convenient access to compute resources. STFC should provide guidance on the use of cloud computing, including ensuring best value and appropriate security.

STFC should provide mechanisms to ensure optimisation of codes so that they make the best use of new hardware. This could be delivered through a mechanism similar to the current Collaborative Computational Projects (CCPs - see p21). Regardless of the mechanism, it is important that experts in new architectures are brought together with those with a detailed knowledge of computing requirements and codes for a specific community.

To promote best practice in the sharing and management of research data, STFC should ensure the *RCUK Common Principles on Data Policy* and *Guidelines on Best Practice in the Management of Research Data* are observed by all data owners using facilities operated or funded by STFC.

Research projects with significant computing requirements should be supported to explore opportunities for industrial engagement.

See the Cross-Programme section for cross-programme recommendations, e.g. on data centres, long-term planning, authentication and authorisation solutions, and code optimisation.

Recommendation 1: STFC should plan for an order of magnitude increase over the next five years in the computing hardware, storage and network bandwidth available to PPAN researchers. The existing hierarchy (from Tier-1 to Tier-3) should be maintained and strengthened whilst ensuring access to a diverse set of compute resources (high-memory, data intensive, high throughput, massively parallel).

Recommendation 2: STFC should ensure that access to Tier-0 scale computing remains available to those PPAN researchers who require it.

Recommendation 3: STFC should make it easier for the PPAN community to access computing resources. Resources should be better publicised and guidance provided to ensure that projects are making use of the most appropriate resources. When required, assistance should be provided to people moving up (or down) the tiers.

Recommendation 4: STFC should ensure the *RCUK Common Principles on Data Policy* and *Guidelines on Best Practice in the Management of Research Data* are observed by all data owners using facilities operated or funded by STFC.

The UK National Facilities

a) Current usage

STFC provides support for computing at the UK national facilities, CLF, Diamond and ISIS, in the following categories: HPC/HTC support; data services (accounting/configuration data management, data storage, back-up and recovery services, databases and repositories, data management); data processing, visualisation and analysis; software development, training and community support; networks; and direct facility operations. The majority of these services are based at RAL, co-located with the facilities, although the Hartree Centre and the UK eScience Certification Authority are based at Daresbury.

The Scientific Computing Application Resource for Facilities (SCARF) cluster provides HPC resource for facility users and STFC staff, for such as molecular simulation work. SCARF is a Tier-2 scale cluster which has several hundred users. CLF operates a data management service that catalogues data and metadata from the Astra-Gemini laser system, enabling it to be displayed in near real-time. ISIS operates data storage software and has a data migration facility to migrate data from disk to tape library and for digital preservation of user data. All ISIS data is held in a data archive at RAL for long term curation.

Both CLF and ISIS have internal database services that support user activities such as proposal management and user accounts. External access to scientific and technical output from ISIS and CLF is provided by an archive with metadata of research publications associated with the facilities, whilst Diamond operates similar archive access. Over 14,000 users have accessed this archive over the last 12 months. Diamond runs a data service, aiming eventually to give off-site access for all end users of the data archive.

Underpinning all these activities is the STFC Network service, allowing transfer of data across the STFC sites and export from the facilities to user groups where necessary.

No current gaps in provision have been identified; however significant trends and increasing requirements for the immediate future are evident due to the acquisition of ever larger data sets.

b) Current plans; long term requirements

Users of all three facilities identified significant changes in their e-infrastructure requirements that are expected to occur over the next five years. There is an increasing need to combine experimental data with theoretical modelling and simulations. Analysis of data sets and modelling of data are already the rate determining steps limiting throughput at the facilities and therefore investment in e-infrastructure will have significant benefits in output and impact from the facilities.

Access to such real-time modelling and simulations will increase the productivity of the facilities by enabling targeted and better designed experiments. Coordinating access to HPC resource with facility access applications across the facilities will be needed to maximise the added value of providing HPC time. In some instances small amounts of HPC resource would be valuable as pump priming to facilitate better experimental design prior to formal applications for facility time. As the algorithms need to become more computationally expensive, there will be a need for significant new facilities for data reduction and modelling, including over the long term when the users will no

longer be on site. High capacity central resources will be needed to process data in real-time, with visualisation capability, and high throughput simulation computing will be needed.

Fast and reliable network connections must be maintained across site and externally, to transfer data and to ensure transparent, fast and accurate access for users to archived data held at the facilities. Large datasets are difficult to transfer to home institutions so options for back-up data sets, (such as those being developed by Diamond [see section a above]) or remote access for users need to be developed further.

In parallel, there is expected to be an increasing need for expert computing support, such as code developers and software support, as relatively few of the facility users will be expert users in terms of computing expertise. Provision of computing support through shared support across the facilities would foster best practice and promote coordination of provision, particularly for users working at more than one facility.

c) Issues and opportunities

The trends driven by the ever increasing data volumes show a growing need for computing provision and support at the facilities to gain maximum impact from the facilities. In some cases these are new costs but in other instances they represent the transfer of activity that was previously carried out at user institutions (such as data storage), to the facilities. It is important to seek the optimum balance between provision on and off-site, recognising that many users will not be experts in computing and will regard this as a support activity they require from the facility and something that they wish to be 'invisible' to them. The increased demand for computing, in terms of hardware and staff support, will likely be reflected in increased operating costs at the facilities; however these are not all necessarily increased costs for the science base.

The skills and expertise identified by the users from STFC focus on expertise in compute, algorithm/ code development and support, and user support for computing. Throughout the facilities the use of common platforms, and simple data formats is preferred, with common interfaces to simulation packages. STFC should take the lead to drive best and common practice across the facilities.

Data storage for the facilities may offer opportunities for economies of scale with other activities across STFC although logistical issues such as security and access would need to be addressed.

d) Recommendations

STFC should consider how access to high performance computing capabilities can be made available to facility users, for pump priming experiments, and for modelling, simulation and visualisation. This could be through expansion of the SCARF facility on site or through collaboration with other providers such as ARCHER (the national HPC service) or the Hartree Centre. Provision of HPC resources should avoid the need for parallel applications to facility access and potential 'double jeopardy' where the HPC access is an integral part of the preparations for the experiment.

Networking provision should be reviewed to ensure transfer across the site is robust and STFC should undertake a detailed review of requirements for the transfer of data to user sites, and for access to archived data at the facilities. STFC should ensure that the requirements for network transfer and computing at the user sites is defined in collaboration with the users to ensure

successful transfer. The need for remote access to be straightforward for non-expert users must be taken into account.

Due to the increasing size of data sets, the need for accessible data storage will likewise increase. Automated systems to store, access and analyse data will be essential. Working with representatives of the user community, STFC should develop a model for the optimum balance between storage on site, at the user institutions and through Cloud based solutions. This model should consider opportunities for sharing of resources such as data storage with non-facility activities within STFC and the user institutions.

STFC should review the computing support (hardware, software and expertise) provided to users, focussing on the support required uniquely at the facilities, promoting coordination and best practice between the facilities. STFC computing support must recognise the differing needs of various research communities and be responsive to the needs of the users, whilst ensuring consistency. Such provision should be included in the operating costs of the facilities.

See the Cross-Programme section for cross-programme recommendations, e.g. on support for users, interoperable software and interfaces, standard data formats and metadata etc.

Recommendation 5: STFC should consider how access to high performance computing capabilities can be made available to facility users, for pump priming experiments, and for modelling, simulation and visualisation. Provision of HPC resources should avoid the need for parallel applications to facility access and potential 'double jeopardy' where the HPC access is an integral part of the preparations for the experiment.

Recommendation 6: Networking provision for the facilities at the RAL site should be reviewed to ensure that the provision keeps pace with capacity requirements and remains robust. The STFC campus network group should be supported to undertake developments not only for internal capacity, but for resilience and to improve the network management structure.

Recommendation 7: STFC should undertake a detailed review of requirements for the transfer of data to facility user sites, and for access to archived data at the facilities, considering the need for remote access to be straightforward for non-expert users.

Recommendation 8: Working with representatives of the facility user community, STFC should develop a model for the optimum balance between storage on site, at the user institutions and through Cloud based solutions. This model should consider opportunities for sharing of resources such as data storage with non-facility activities within STFC and the user institutions.

Engineering Support

a) Current usage

As well as operations and data analysis, the design and construction of STFC Facilities and experiments within the PPAN programme is already, and increasingly, heavily dependent upon access to advanced simulation, Computer Aided Design (CAD), Computer Aided Engineering (CAE) and Computer Aided Manufacturing (CAM). This is true across a wide range of engineering disciplines, from optics, through mechanical engineering and fluid dynamics, to electronics design, device simulations and satellite control.

Much of the engineering design and simulation work is carried out with sophisticated commercial packages that increasingly integrate design and simulation. They are demanding applications, often with strict licensing restrictions and requiring high-end dedicated Windows or Unix workstations[1]. However, there is also overlap with other aspects of computing such as, for example, the use of GEANT for modelling radiation effects, and the computational fluid dynamics developments of CCP12. The applications themselves are increasingly demanding in terms of computing power and data output. The nature of the broader market for the commercial products, however, tends to focus them on high-end dedicated workstations rather than shared HPC systems. As the simulation demands grow, we might expect to see this distinction blurred in the future.

STFC supports access to advanced engineering primarily, though not exclusively, through its Laboratory sites. Sharing across sites is often restricted by license conditions. In some areas, such as micro-electronics, the laboratories also provide access to and support for advanced CAD tools for academic users in universities and research institutes. This includes the training for some of the microelectronics tools (e.g. FPGA programming). In this case STFC must meet stringent security requirements set by the software vendors. Failure to do this would jeopardise access to these advanced tools for HEI and laboratory staff.

Underpinning all these activities is the STFC Network service, allowing transfer of data across the STFC sites and export from the facilities to user groups where necessary.

No current gaps in provision have been identified beyond the general trend towards more provision driven by greater integration, more sophisticated simulation, and increasing data sizes.

b) Current plans; long term requirements

Over the next 5 years there will be a gradual increase in computing requirements, probably matched to the evolution of high-end windows/linux servers, although the increasing demands of ever more sophisticated simulation and modelling may open up new requirements for access to HPC machines.

The current trend is for increased integration of design and modelling, leading to increased productivity. In micro-electronics the increasing complexity and cost of manufacture is driving a greater focus on simulation and verification before committing to costly fabrication runs. In

[1] These different operating systems are not necessarily interchangeable; e.g. in the micro-electronics industry Unix is more important than Windows.

mechanical engineering, the growth of additive manufacturing (3-d printing) will also see the manufacturing itself increasingly integrated into the design process.

In some areas, such as microelectronics and detector system design, there is an emerging trend towards further restricting access to software, requiring users to use remotely hosted services and data libraries, without direct access to the software. As trusted partners with a track record in supporting the academic community the National Laboratories are well placed to be able to offer these services to this community in situations where the software providers themselves do not wish to do so. This would be a means of maintaining UK academic access to advanced tools. If this trend continues it will naturally lead to increased demand for "cloud" like services such as those already provided by STFC.

In parallel, there is expected to be an increasing need for expert computing support, as packages become more complex to operate and administer, fewer users will be expert users in terms of computing expertise. Provision of computing support through shared support across the facilities/departments/institutes would foster best practice and promote coordination of provision.

Over recent years it has also been increasingly difficult to secure academic licenses for many products used within STFC Laboratories. This trend is likely to continue given the greater emphasis on industrial collaboration on the STFC campuses. Therefore, it will be important to make effective provision for the cost of commercial licenses, and to ensure efficient and effective use of licensed software.

c) Issues and opportunities

Many of the issues faced are common to the STFC facilities exploitation described elsewhere:

- The increased demand for computing, in terms of hardware and staff support
- Ever increasing data volumes and the data management tools required to support this
- Restrictive licensing often not well matched to distributed academic collaborations

Access to HPC systems and dedicated emulators for hardware modelling and software development may be required in the near future.

In some cases it is quite possible that the overall cost of ownership of the most sophisticated tools (including installation and management of advanced design tools), coupled with the security requirements (design tools and design data) will become so high (particularly in the microelectronic sector) as to be impractical for most, if not all, academic groups. It is possible that STFC National Laboratories can provide the critical mass required to offer a managed service to the academic community and their collaborators. This model is effectively already in operation in some areas such as calculating orbits of spacecraft, where the tool suites are sufficiently complex that consortium partners often look to STFC to do these calculations.

d) Recommendations

STFC should consider how access to high performance computing capabilities can be made available to engineers and technical staff for modelling, simulation and visualisation. In practice uptake will be dependent upon the relevant software availability for the platforms in question.

Networking provision should be reviewed to ensure that software can be accessed effectively and that data transfer across the site is robust. Due to the increasing size of data sets, the need for accessible data storage will likewise increase.

STFC computing support must recognise the differing needs of various engineering communities, requiring access to best in class commercial software as well as providing externally facing support services in microelectronics that need to meet stringent requirements in authentication and authorisation set by the software vendors themselves.

STFC should explore opportunities/requirements for managed[2] services to the academic community in order to access the most sophisticated tools.

For networks see Recommendation 6.

Recommendation 9: STFC should continue to make provision for the cost of engineering CAD/CAM/CAE licenses, both commercial and non-commercial, and to ensure efficient and effective use of licensed software.

Recommendation 10: STFC should encourage sharing of expertise, support, and, where possible, licenses for commercial engineering software tools.

Recommendation 11: STFC should explore opportunities/requirements for managed services to the academic community in order to access the most sophisticated engineering software tools.

[2] In this context, "managed" is assumed to include training/expert support in exploiting the tools in question.

Externally Funded Projects and Services

As well as providing computing for STFC programmes in Particle Physics, Astronomy and Nuclear Physics and the National Laboratories, STFC also provides computing for many other areas of the UK research base through services and joint research with academic and industrial organisations. Customers (and funders) for these services include other Research Councils, the European Commission (EC), universities and commercial clients. The nature of these services is very varied, but many make use of expertise within the Scientific Computing Department (SCD) and contribute to the STFC mission by delivering world-class research, innovation and skills in scientific computing.

a) Current usage

Current externally funded projects and services includes, for example:

- Collaborative Computational Projects
- JASMIN/Climate and Environmental Monitoring from Space Facility (CEMS)
- Emerald
- EC Projects and other grant funded work
- National Service for Computational Chemistry Software (NSCCS)
- Medical Research Council (MRC) Data Support Service
- Hartree Centre– see next section
- Micro-electronics Support Centre (not using SCD)
- British Atmospheric Data Centre

There are also a number of other similar collaborations under development including with the Culham Centre for Fusion Energy (CCFE), the Genome Analysis Centre, the European Centre for Medium Range Weatherforecasting and with individual universities.

Collaborative Computational Projects

The CCPs are software development projects that bring together researchers and developers in particular topics in computational science. By co-ordinating the work of many groups, the CCPs deliver leading research and high quality sustainable software. They also promote collaboration with industry. The CCPs are funded by the EPSRC, MRC and BBSRC and the US National Science Foundation and delivered and managed by the STFC SCD.

JASMIN/CEMS

The JASMIN facility is a "super-data-cluster" which delivers infrastructure for analysis of environmental data. It provides a globally unique computational environment that includes a high performance computing resource and a private cloud, both configured with access to a petascale data storage centre. CEMS is the Earth Observation activity which runs on the JASMIN infrastructure, alongside other services provided by JASMIN. JASMIN is funded by the Natural Environment Research Council (NERC) and delivered by the STFC SCD.

Emerald

Emerald is a GPU cluster that is run by STFC SCD as a service for STFC and Oxford, University College London, Bristol and Southampton universities. When installed it was a first, being the largest GPU cluster in the UK, and its success has led to other GPU clusters being installed at universities, including at some of the partner universities . The Emerald hardware is now reaching end-of-life however the equipment is being refreshed, albeit at a smaller scale, and the service will continue for use by STFC and some of the university partners.

National Service for Computational Chemistry

The NSCCS is an EPSRC funded service led by Imperial College London and delivered in part by the SCD. STFC provides hosting and user support at a higher level than can be expected from an individual university. Hardware could be shared with other services provided the NSCCS computation library remains available.

Micro-electronics Support Centre

The Microelectronics Support Centre (MSC) is dedicated to stimulating the use of modern electronic design techniques and tools in academic institutions throughout Europe to enhance their microelectronic and microsystem design, research and training capabilities. This is achieved by offering a comprehensive range of services including technical consultancy, access to industry-leading electronic design tools and dedicated training courses to more than 600 academic institutions across Europe. The MSC operates the EUROPRACTICE Software Service. EUROPRACTICE is an EC initiative which aims to stimulate the wider adoption of state-of-the-art microelectronics and electronic system design methodologies and technologies within Europe. The EUROPRACTICE Software Service provides easy access to a wide range of cost-effective leading-edge system design tools to European academic institutions and publicly funded research laboratories for their non-commercial teaching and research.

Projects for other Research Councils

The MRC Data Support Service facilitates data sharing for several MRC population and patient studies. It provides a catalogue of datasets of 35 of the largest MRC-funded and co-funded cohort studies (see https://www.datagateway.mrc.ac.uk/). The metadata is provided down to variable level in order to optimise the long-term use of these data assets. STFC contributed to the design and development of the system and definition of a metadata model. The service is run by STFC SCD on behalf of MRC.

The Biotechnology and Biological Sciences Research Council (BBSRC) Data Service provides a data backup service to BBSRC institutes based upon technology developed for the LHC Tier1.

b) Current plans; long-term requirements

Across these services a number of common areas for development arise. These are described below.

The greater digitisation of science, together with new high resolution instruments, and more powerful simulations requires substantial increases in compute, data storage, data transfer and data management capacity. This will require access to high-performance, reliable networks; direct access

to data sets for analysis to avoid bottlenecks; and achieving the right mix of high bandwidth storage and computing. Long-term digital curation is a new requirement.

Software development across distributed teams and expansion in software development techniques are seen as essential in several areas. Evolving hardware, particularly the increasing numbers of cores and different processor architectures, will require future-proofing applications to run on various platforms and including optimisation in compilation.

Bringing together domain expertise with specialist computing expertise is critical, for example to make web interfaces interact with jobs on remote resources. High quality operational skills in SCD need to be combined with the ability to innovate.

Access to resources needs to be standardised, for example through common authentication and authorisation mechanisms and the provision of internal private cloud resource. There is a need to prepare for the collision of batch computing, massive storage and virtualisation software in e-infrastructures.

Research and development expertise needs to be maintained in key areas including cloud computing and virtualisation.

c) Issues and opportunities

Many of the issues described above are not unique to any particular domain and STFC should aggregate its experience across domains and deliver facilities and services which exploit this cross-domain knowledge. For some of these common topics, such as solutions at scale, exploiting novel architectures and combination of simulation and experimentation, requirements are pushing the envelope of technology and require a Research and Development (R&D) programme to ensure technological development is fit for purpose.

Services such as JASMIN and Emerald have led the way combining hosting and high quality system support with deep technical expertise. These innovations provide a basis for rollout to other projects. For example it may be possible to apply the JASMIN model to other major projects such as SKA where SCD experience with mixed cloud and batch computing is being seen as an exemplar. Continued participation in leading projects such as these will be required, particularly around cloud computing and virtualisation, in order to continue to innovate in these areas into the future.

Discussions with organisations providing services to other domains, such as CCFE, suggest that expertise developed for the STFC programme could add value to other areas. The value to be added is in bringing highly specialised experts in computing domains (data, compute, networking, software for scientific applications) together with experts from the scientific domains. Within STFC, the SCD is in a unique position in terms of scale and track record or delivering scientific computing infrastructure, technology and applications.

Although many services remain highly bespoke and dedicated to a single science domain, the user community served by these services is often very large. However, opportunities for further consolidation exist. JASMIN is a good example of national consolidation within a discipline; further consolidation across disciplines is possible especially at lower levels of the software stack. This is probably the area where groups and disciplines without access to their own dedicated computing infrastructure should be encouraged to focus.

Facilities like JASMIN and Diamond exemplify a trend that storage bandwidth is not growing as fast as storage capacity, and capacity is not growing as fast as data production. These trends are increasingly leading to requirements for the data to be held centrally and for compute to be near the data. Furthermore, this concentration of resources requires a heterogeneous software environment and significant expertise throughout the computing stack, for example in high volume data, extremely large file systems, HPC-like computing etc.

d) Recommendations

STFC should develop criteria relating to the added and unique value that its expertise can provide, to use in deciding whether to host and operate facilities for other partners.

Experience with facilities such as JASMIN and Emerald has shown that these can be provided efficiently by building on the expertise and infrastructure developed for STFC's core programme. Opportunities to expand this model for delivery of large scale data facilities to other areas of science should be sought and evaluated.

The organisational co-location of expertise in computational science, experimental data management and analysis, and high performance computing, in particular in novel architectures, provides an opportunity to develop collaborations with the breadth of technical knowledge to tackle major scientific problems in new ways. STFC's Scientific Computing Department currently provides services for five of the seven Research Councils (BBSRC, EPSRC, NERC, MRC, STFC) many of these in collaboration with university partners. This "hub" of computing resource and expertise is a significant national asset that demonstrates the value of collaboration between Research Councils in providing multidisciplinary facilities and technologies such as computing. STFC should ensure that it maintains the skills and capabilities needed to support large-scale scientific computing facilities and collaborates with other centres of expertise such as the EPCC.

Recommendation 12: STFC should develop criteria to use in deciding whether to host and operate facilities for other partners. These should assess the added and unique value that its expertise can provide.

Recommendation 13: Opportunities to build on the expertise and infrastructure developed to deliver large-scale data facilities for STFC's core programme and apply the models to other areas of science should be sought and evaluated.

Recommendation 14: STFC should work with the other Research Councils and other centres of computing expertise, such as the EPCC, to identify where STFC can contribute to their research programmes by providing expertise and supporting services.

Recommendation 15: STFC should ensure that it maintains the skills and capabilities to support large-scale scientific computing facilities that are an essential requirement for a broad UK science base.

Hartree Centre

The Hartree Centre was established in 2012. It is based at Daresbury, and provides HPC systems for modelling, simulation, visualisation, big data analytics and energy efficient computing. Its mission is primarily to work in collaboration with industry and business partners to bring them competitive advantage through access to these tools and systems. It contributes significantly to the development of the Daresbury Campus and as an attractor to investment. The Department for Business, Innovation and Skills (BIS) provided £57m of capital investment, recently supplemented by £115m to expand the data centric and cognitive computing capabilities, and the Hartree Centre requires approx. £5m of operating costs per year. Initially, £3m per year was provided by the Business Innovation Department of STFC. The resource funding from STFC is reducing year on year as the proportion of funding generated by external income increases. Hartree has around 1000 registered users.

a) Current usage

The current service model employed by the Hartree Centre to provide capability to UK industry and academia to use these technologies has five service offerings:

- You solve the problem – Access to hardware (petascale HPC, HTC or storage)
- We solve the problem – Consultancy/professional services
- We develop a solution – Software development
- We solve the problem together – Collaborative R&D
- We train you to solve the problem – Training & Education

Most of the offerings to UK academia fall into the first and last of these categories. The hardware is not operated as a service facility, although spare HPC capacity is made available to academic groups where feasible on a best efforts basis. Hartree provides expertise in code optimisation for new and emerging technologies, expertise in the use of existing modelling packages used by communities, and domain specific computing expertise, e.g. computational modelling around engineering or fluid dynamics.

b) Current plans; long-term requirements

With the recent announcement of additional capital funding by BIS, it is planned to expand the existing Hartree Centre to provide a joint (STFC/IBM) research facility embedded in Hartree Phase 3. This research centre will look at a number of aspects of data centric and cognitive computing, and has e-infrastructure requirements which underpin it. These requirements include the availability of up-to-date IBM power infrastructure, elastic storage, software and capability to enable sophisticated data manipulation, analytics, and cognitive computing solutions. Hartree Phase 3, an evolution of the Phase 2 work underway currently, is expected to start in Q3 2015.

c) Issues and opportunities

The issues that the Hartree is aiming to address (around future proofing codes for new architectures) affect both industrial and academic users alike.

The Hartree Centre has a distinct role in working with business and the research community in a collaborative manner to develop new tools and systems. Whilst it does not operate as a service facility, spare HPC capacity is made available to the academic community when possible. It is important to work with EPSRC, who has responsibility for the UK national service, ARCHER, to communicate these complementary roles clearly to the academic and industrial communities. Sharing a work programme or workload could produce efficiencies for both ARCHER (and its replacement) and the Hartree Centre, as well as DiRAC and STFC Tier 1.

There is also an obligation on STFC to ensure that, consistent with its mission, any spare capacity is used to the maximum benefit of the UK community. There may be opportunities around the Hartree making its services (and cycles) more widely available to UK academia and the STFC Facilities, without impinging on the Hartree's mission to support industry.

Phase 3 developments remain a significant opportunity, although as this programme is in its early stages, these are yet to be clearly defined.

In a five-year timescale, Hartree may anticipate a number of additional e-infrastructure technologies to embrace. These include 3D printing, with scientists and technicians wishing to design tools and products. This will be data driven with increasingly sophisticated integration of data (including Big Data - both unstructured and structured) from some of the UK's national utilities and services e.g. integration across transport services, banking systems, some aspects of the NHS. Data will need to be moved across the boundaries of these organisations and systems, aggregated and analysed to deliver new insight, services and outcomes, and it is not clear how this will be done. Potentially STFC's role could be in developing best practice to support other Research Council communities to use their data in similar ways, and/or to host and manage data on their behalves.

d) Recommendations

STFC needs to be clear on the future role of the Hartree and where it fits into the overall corporate STFC strategy, recognising its role in the development of the Daresbury Campus. This is a new type of venture for STFC and both its role and fit into the governance of STFC requires development.

STFC should work with EPSRC to clarify and communicate their respective roles in the provision of HPC for UK academia and industry. STFC should be clear about the Hartree position with respect to delivery and support to UK academia.

To maximise benefit to the UK from the Hartree Centre, there should be structured links to potential partners and stakeholders such as the other UK Research Councils. STFC should consider how it might exploit the growing relationship with IBM Research to benefit themselves and the UK Research Councils. STFC is in a unique position with respect to all of the Research Councils as its expertise and skills can underpin delivery of their data intensive science and can drive benefit across this landscape.

Recommendation 16: As the future development of the Hartree Centre will be of critical importance for the Daresbury Campus, industry and government, STFC should ensure that the Hartree Centre's governance structure and its place in STFC corporate strategy are communicated clearly to all stakeholders.

Recommendation 17: STFC should work with EPSRC to clarify and communicate their respective roles in the provision of HPC for UK academia and industry, with respect to DiRAC, the Hartree Centre and ARCHER.

Recommendation 18: To maximise benefit to the UK from the Hartree Centre, there should be structured links to potential partners and stakeholders such as the other UK Research Councils.

Recommendation 19: STFC should ensure the Hartree Centre's position with respect to delivery and support to UK academia is communicated clearly.

Cross-Programme Topics

The following issues have been identified as relevant across more than one programme area.

Long-term planning

Computing is now an essential component of all research projects and can account for a substantial proportion of the project costs. Projects can take years or even decades to develop. Long-term planning for computing requirements is therefore crucial. For the STFC-funded research projects there is a need to plan for computing from the earliest stages of project planning, recognising that detailed plans will not be possible at the grant application stage and the plan will require amending as the project requirements are refined, technology develops and costs change. Timely updates from projects will assist STFC in making the best use of resources. New projects should be encouraged to seek advice from more advanced projects that already employ long-term cost modelling around computing needs. For the facilities, there is a need to plan what user compute support (e.g. data archiving, modelling capability, access to computing power) will be available from (or through) STFC to enable the costs to be incorporated into the operating costs of the facilities and for the user to plan for capabilities required at their own institution.

Recommendation 20: STFC-funded R&D or construction projects should be asked to provide a plan for computing from the earliest stages of project planning and provide STFC with updates as plans and costs develop during the project's lifetime. This is intended so that STFC can have a coherent picture of future expectations for planning purposes.

Recommendation 21: New research projects should be expected to seek advice from more advanced projects that already employ long-term cost modelling around computing needs and, if appropriate, develop partnerships to make use of shared resources.

Recommendation 22: STFC should review what compute support will be required by facility users (e.g. data archiving, modelling capability, access to computing power), focussing on the support required uniquely at the facilities, so it can plan what will be provided by STFC and what users will need to seek elsewhere. This will support STFC's financial planning, by enabling compute support costs to be incorporated in the operating costs of the facilities; and will enable users to plan for capabilities required at their own institution.

Increasing data volume challenges

The huge increase in data volumes anticipated across all areas of STFC computing is one of the biggest challenges for the future. Long-term planning will be essential (see above). There will be an increasing need to move compute provision next to the data to avoid transferring vast quantities of data.

Recommendation 23: STFC and its expert user communities (for example in grid and cloud) should continue and develop their technical leadership and guidance on the challenge of managing increasing volumes of data by encouraging users to adopt architectures that are best equipped to deal with the problem.

Data centres

Several forthcoming STFC research projects have anticipated the challenges of increased data volumes by including requirements for data centres in their plans. Projects should be encouraged to plan for data requirements at the earliest stage of project planning; to share best practice and expertise; and collaborate where appropriate to share resources. Good communication should be maintained with partner organisations such as the UK Space Agency and relevant international organisations and funding agencies on future projects and their computing requirements.

There is also the separate issue of STFC hosting data centres on its campus sites, whether for STFC's own needs, that of its funded research communities/projects or funded externally. STFC needs to develop criteria that will enable it to establish priorities for hosting data centres and work in collaboration with other potential providers such as Jisc. This must be consistent with the needs of all STFC sites, the funded projects and ensure appropriate interaction with any Campus developments around data centres at RAL and Daresbury Laboratory. The establishment of new centres, such as the University based Alan Turing Institute and the Hartree Centre at Daresbury, recognises the need for increased investment and specialisation in data science and analytics. It is essential that synergies and effective modes of collaboration are found between these new centres and with the established national computing centres.

Recommendation 24: STFC should take a more active role in co-ordinating use of e-infrastructure and investigating the opportunity to share resources, without sacrificing the variety and specialist resources required by various communities. New projects should be strongly encouraged to make use of existing expertise and infrastructure. Good communication should be maintained with partner organisations such as the UK Space Agency on future projects and their computing requirements.

Recommendation 25: STFC needs to develop criteria that will enable it to establish priorities for hosting data centres.

Recommendation 26: Synergies and effective modes of collaboration need to be found between new initiatives such as the Alan Turing Institute and Hartree Centre and the established national computing centres.

Cloud computing/Virtualisation

Increasingly users wish to submit compute jobs without needing to be concerned about where they will be carried out. The adoption of cloud computing is growing with various models available including public, private and hybrid clouds. However, there is a lack of guidance on how best to utilise and access cloud compute resources. STFC should also work towards having a unified method of accessing its dedicated hardware, and a straightforward way to move computing between different resources. STFC needs to ensure that compute resources are delivered for its users in the most efficient way. STFC has the expertise in cloud computing to provide technical leadership in this area. STFC should engage with other Research Councils and the National Project Directors Group who are active in this field.

Networks

Networks are an underpinning technology, essential to all areas of STFC computing. This includes networks at STFC laboratory sites (the responsibility of STFC's Corporate Information and Communications Technology department) and university links to the Janet network. In the future some major research projects, such as LSST and SKA, anticipate the need for new, dedicated links. There is a growing requirement for data to be transferred from the Tier 1 data centres and compute resources at the facilities to universities so that users can carry out complex post-processing visualisation locally. This may require improvements to networks, including the "last mile" across campuses. STFC must ensure good communication with Jisc continues to plan for the networking capability that will be required.

Recommendation 27: STFC must communicate the community's needs to Jisc, to ensure that Janet continues to provide the network capacity required to support STFC computing in the UK. In addition, links to European and worldwide networks, for example through GEANT, must be maintained as these are vital for international projects.

Career paths for data scientists

There is an emerging need for data scientists in research environments, to work as an interface between the scientists and computing hardware experts. At present, computing does not have as clearly defined a career path as the standard academic or managerial track. These types of posts are vulnerable to the short-term nature of grant funding compared to the long-term need for these posts on research projects. In general, STFC should ensure it is aware of emerging skills gaps in computing that could affect its programmes and consider initiatives to address any such gaps.

Recommendation 28: STFC should ensure that the career path of its professional computing staff is appropriate, and equivalent to that of comparable professionals within STFC. STFC should encourage Universities to support the long-term development of computing skills and establish a career path for software engineers and data scientists.

Recommendation 29: STFC needs to provide guidance to grants panels on the role for software engineers and data scientists in research projects and how to handle funding requests for such posts.

Support for non-experts

Users of STFC computing need appropriate knowledge and expertise in order to make the most effective use of the resources. In many cases, it is impractical (and undesirable) for them to become experts in algorithm development, large scale computing and code optimisation. This expertise should be provided through guidance on best practice, training or, for advanced skills such as algorithm/code development, access to skilled support staff. The provision of these resources should be factored into the costs of STFC e-Infrastructure. At the facilities, common software packages and interfaces would be beneficial as users often use more than one facility, so facility unique solutions should be avoided unless essential.

Recommendation 30: STFC should ensure that users are provided with appropriate support, whether through best practice guidance, training or access to skilled support staff, acknowledging that it may be impractical (and undesirable) for users of STFC computing to have advanced computing skills. These costs should be factored into the costs of STFC e-infrastructure. STFC should support a community of expertise, whether at universities or STFC Facilities.

Recommendation 31: As users often use more than one facility, STFC should promote co-ordination and best practice in its computing support. Interoperable software and interfaces should be used where possible; standard data formats and metadata should be encouraged.

Increasing requirements for modelling/simulation

This is a strongly growing trend in many disciplines. For example, order of magnitude increases in modelling/simulation are anticipated for many areas within the particle physics, astronomy and nuclear physics programmes, keeping pace with expected increases in data processing. One area of expansion is the increasing desire to obtain statistical uncertainties on simulation results by re-running them numerous times. There is also a growing move to carrying out simulations alongside experiments. Facility users have an increasing need to carry out real-time modelling or simulation or preliminary modelling work to enable experiment design. This will increase facility productivity by enabling targeted and better designed experiments. Various options exist to support the increased need for modelling/simulation, for example via grant funding or direct funding from STFC.

Image processing, visualisation and analysis

There is a trend towards visualisation being distributed, remote from the compute. There is a strong need in the facilities for near real-time analysis. This improves the efficiency of experiments as understanding if the experiment is working or not allows parameters to be changed on site instead of having to wait until the user is back at their host institute. Near real-time analysis is also increasingly important in astronomy and will be required for future astronomy projects such as SKA and LSST.

Authentication and authorisation

Authentication and authorisation solutions will facilitate access to compute resources, for example standardised systems would facilitate the sharing of resources such as DiRAC and GridPP or the sharing of data amongst international collaborations. Facility users require remote as well as on-site access to facility data archives. The remote access needs to be simple and easy for users to achieve. This is a wider issue than STFC and some initiatives are underway nationally to address it.

Recommendation 32: STFC should support solutions into authentication and authorisation that would have wide applications for STFC-supported research communities. STFC should engage with other research councils and also international collaborations where feasible that are carrying out work in this area.

Software – sustainability and portability

Software is vital for scientific research, but as a sometimes significant cost, users need guidance to maximise investment in the long-term, since code development and lifetimes often have to evolve

through many generations of hardware technology. Changes in computing architectures will require significant effort to port codes to new platforms. STFC should encourage code optimisation where appropriate to make the best and most cost-effective use of hardware. Sometimes a relatively small investment in software development can lead to a significant reduction in capital hardware spend. Ways to support innovative software development should be considered, as techniques need to keep pace with scientific change. Models such as the CCPs could be considered for areas of PPAN science.

Recommendation 33: STFC should provide mechanisms to ensure optimisation of codes to make the best use of new hardware. Possible models to support innovative software development should be investigated.

Foresight on emerging developments in compute technology

Developments in compute architectures will affect all compute users. STFC's SCD have the expertise to lead on this and provide guidance across STFC programmes to maximise the benefits and reduce the risks, although this may require additional resource. There is also significant expertise in the universities that could play a major role in this.

Recommendation 34: STFC should make full use of its technical expertise, whether in the Scientific Computing Department or universities, to provide guidance across STFC programmes on emerging developments in compute technology to maximise the benefits and reduce the risks for users of STFC computing.

Cloud computing (see above) is one area where STFC could provide technical leadership and guidance. Another is the technical issues now around hitting ceilings on data transfer rates to and from disc (and tape) which leads to increasingly long buffering times, while more data needs to be written or read. This problem will increase as data volumes increase and disc technology fails to address it. Moving compute to the data will not help. Maintaining close relationships with vendors (see below) should help to better understand their potential future solutions.

International relationships

STFC's international relationships on computing include international research bodies such as CERN or ESO; partners in EC funded programmes (e.g. Horizon 2020 collaborations; the EUROPRACTICE Software Service); infrastructure providers (e.g. Geant, PRACE); international collaborations to deliver the largest scale scientific projects and their associated computing (e.g. the Worldwide LHC Computing Grid) and other initiatives aiming to co-operate on computing (e.g. the EUT0 federation of European funding agencies and CERN). Engaging with international partners is essential to deliver or secure access to the largest scale scientific or compute facilities; shape future capability; share resources; achieve economies of scale and share best practice and expertise.

Relationships with vendors

Maintaining relationships with vendors will enable early knowledge of emerging technologies and facilitate planning. Collaborations with vendors on development projects (e.g. DiRAC) can provide STFC funded scientists with early access to cutting-edge technology.

Operating Costs

Large compute facilities require major capital investment, but also have significant on-going operating costs. Historically, Government funding for compute facilities has tended to provide the capital costs without the associated recurrent costs. If future funding opportunities for computing infrastructure become available, the need to include recurrent costs should be emphasised.

Large computing facilities have very high energy consumption and this constitutes a large fraction of the operating costs. Existing and future facilities should be encouraged to find the most cost-effective solutions in terms of selecting energy suppliers and reducing energy consumption. Data centres should be encouraged to sign up to the *European Code of Conduct for Energy Efficiency in Data Centres*.

Procurement

Within constraints of timescales and other project dependencies, projects should consider the feasibility of undertaking joint procurements to maximise value. Such discussions may also encourage projects to work together to share computing infrastructure.

Recommendation 35: STFC projects should be encouraged to consider the feasibility of undertaking joint procurements in computing to maximise value and where possible use recommended configurations.

Recommendations

Particle Physics, Astronomy and Nuclear Physics Research Programme

1. STFC should plan for an order of magnitude increase over the next five years in the computing hardware, storage and network bandwidth available to PPAN researchers. The existing hierarchy (from Tier-1 to Tier-3) should be maintained and strengthened whilst ensuring access to a diverse set of compute resources (high-memory, data intensive, high throughput, massively parallel).

2. STFC should ensure that access to Tier-0 scale computing remains available to those PPAN researchers who require it.

3. STFC should make it easier for the PPAN community to access computing resources. Resources should be better publicised and guidance provided to ensure that projects are making use of the most appropriate resources. When required, assistance should be provided to people moving up (or down) the tiers.

20. STFC-funded R&D or construction projects should be asked to provide a plan for computing from the earliest stages of project planning and provide STFC with updates as plans and costs develop during the project's lifetime. This is intended so that STFC can have a coherent picture of future expectations for planning purposes.

21. New research projects should be expected to seek advice from more advanced projects that already employ long-term cost modelling around computing needs and, if appropriate, develop partnerships to make use of shared resources.

29. STFC needs to provide guidance to grants panels on the role for software engineers and data scientists in research projects and how to handle funding requests for such posts.

Facilities

5. STFC should consider how access to high performance computing capabilities can be made available to facility users, for pump priming experiments, and for modelling, simulation and visualisation. Provision of HPC resources should avoid the need for parallel applications to facility access and potential 'double jeopardy' where the HPC access is an integral part of the preparations for the experiment.

6. Networking provision for the facilities at the RAL site should be reviewed to ensure that the provision keeps pace with capacity requirements and remains robust. The STFC campus network group should be supported to undertake developments not only for internal capacity, but for resilience and to improve the network management structure.

7. STFC should undertake a detailed review of requirements for the transfer of data to facility user sites, and for access to archived data at the facilities, considering the need for remote access to be straightforward for non-expert users.

8. Working with representatives of the facility user community, STFC should develop a model for the optimum balance between storage on site, at the user institutions and through Cloud based solutions. This model should consider opportunities for sharing of resources such as data storage with non-facility activities within STFC and the user institutions.

22. STFC should review what compute support will be required by facility users (e.g. data archiving, modelling capability, access to computing power), focussing on the support required uniquely at the facilities, so it can plan what will be provided by STFC and what users will need to seek elsewhere. This will support STFC's financial planning, by enabling compute support costs to be incorporated in the operating costs of the facilities; and will enable users to plan for capabilities required at their own institution.

Engineering Support

9. STFC should continue to make provision for the cost of engineering CAD/CAM/CAE licenses, both commercial and non-commercial, and to ensure efficient and effective use of licensed software.

10. STFC should encourage sharing of expertise, support, and, where possible, licenses for commercial engineering software tools.

11. STFC should explore opportunities/requirements for managed services to the academic community in order to access the most sophisticated engineering software tools.

Externally Funded Projects

12. STFC should develop criteria to use in deciding whether to host and operate facilities for other partners. These should assess the added and unique value that its expertise can provide.

14. STFC should work with the other Research Councils and other centres of computing expertise, such as the EPCC, to identify where STFC can contribute to their research programmes by providing expertise and supporting services.

Hartree Centre

16. As the future development of the Hartree Centre will be of critical importance for the Daresbury Campus, industry and government, STFC should ensure that the Hartree Centre's governance structure and its place in STFC corporate strategy are communicated clearly to all stakeholders.

17. STFC should work with EPSRC to clarify and communicate their respective roles in the provision of HPC for UK academia and industry, with respect to DiRAC, the Hartree Centre and ARCHER.

18. To maximise benefit to the UK from the Hartree Centre, there should be structured links to potential partners and stakeholders such as the other UK Research Councils.

19. STFC should ensure the Hartree Centre's position with respect to delivery and support to UK academia is communicated clearly.

26. Synergies and effective modes of collaboration need to be found between new computing initiatives such as the Alan Turing Institute and Hartree Centre and the established national computing centres.

Cross-Programme Topics

Providing guidance and technical leadership

4. STFC should ensure the *RCUK Common Principles on Data Policy* and *Guidelines on Best Practice in the Management of Research Data* are observed by all data owners using facilities operated or funded by STFC.

23. STFC and its expert user communities (for example in grid and cloud) should continue and develop their technical leadership and guidance on the challenge of managing increasing volumes of data by encouraging users to adopt architectures that are best equipped to deal with the problem.

34. STFC should make full use of its technical expertise, whether in the Scientific Computing Department or universities, to provide guidance across STFC programmes on emerging developments in compute technology to maximise the benefits and reduce the risks for users of STFC computing.

Improving effectiveness

13. Opportunities to build on the expertise and infrastructure developed to deliver large-scale data facilities for STFC's core programme and apply the models to other areas of science should be sought and evaluated.

15. STFC should ensure that it maintains the skills and capabilities to support large-scale scientific computing facilities that are an essential requirement for a broad UK science base.

24. STFC should take a more active role in co-ordinating use of e-infrastructure and investigating the opportunity to share resources, without sacrificing the variety and specialist resources required by various communities. New projects should be strongly encouraged to make use of existing expertise and infrastructure. Good communication should be maintained with partner organisations such as the UK Space Agency on future projects and their computing requirements.

25. STFC needs to develop criteria that will enable it to establish priorities for hosting data centres.

27. STFC must communicate the community's needs to Jisc, to ensure that Janet continues to provide the network capacity required to support STFC computing in the UK. In addition, links to European and worldwide networks, for example through GEANT, must be maintained as these are vital for international projects.

28. STFC should ensure that the career path of its professional computing staff is appropriate, and equivalent to that of comparable professionals within STFC. STFC should encourage Universities to support the long-term development of computing skills and establish a career path for software engineers and data scientists.

30. STFC should ensure that users are provided with appropriate support, whether through best practice guidance, training or access to skilled support staff, acknowledging that it may be impractical (and undesirable) for users of STFC computing to have advanced computing skills. These costs should be factored into the costs of STFC e-infrastructure. STFC should support a community of expertise, whether at universities or STFC Facilities.

31. As users often use more than one facility, STFC should promote co-ordination and best practice in its computing support. Interoperable software and interfaces should be used where possible; standard data formats and metadata should be encouraged.

32. STFC should support investigations into authentication and authorisation solutions that would have wide applications for STFC-supported research communities. STFC should engage with other research councils and also international collaborations where feasible that are carrying out work in this area.

33. STFC should provide mechanisms to ensure optimisation of codes to make the best use of new hardware. Possible models to support innovative software development should be investigated.

35. STFC projects should be encouraged to consider the feasibility of undertaking joint procurements in computing to maximise value and where possible use recommended configurations.

Annex 1: The Strategic Review Development Process

Computing Strategic Review Working Group

The Strategic Review development was led by a working group comprising:

- Juan Bicarregui, Scientific Computing Department
- David Corney, Scientific Computing Department
- Stephen Fairhurst, Chair of the Computing Advisory Panel
- Rachel Reynolds, Programmes
- Janet Seed, Programmes

Survey of current computing facilities and services

The working group carried out a survey in November 2014 of STFC's current computing facilities and services. The survey included questions about the purpose of the facilities and services; location; year started/commissioned; any STFC capital or recurrent funding; types and numbers of users; and the skills required to run them. The facilities and services included those in support of STFC programmes, National Facilities or campus centres as well as those provided by STFC to external organisations. 76 responses were received, including 13 for various different GridPP Tier 2 university sites and 16 covering the different components of CLF computing. Summaries of the survey responses can be found in Annex 2. The survey responses were used to inform the descriptions of 'current usage' in the Review sections on the different programme areas.

Consultation on future computing requirements and opportunities

The consultation on future computing requirements and opportunities to improve STFC's computing provision was developed with the assistance of a Consultation Steering Group. This group contained representatives of the PPAN community, Facility Directors, Diamond and the RCUK E-Infrastructure Group. The steering group advised on the consultation questions and the list of stakeholders to be included in the consultation. The consultation questions were divided into specific questions about the future e-infrastructure requirements of current and planned projects, facilities and services as well as more general questions such as what opportunities there are for STFC to improve the provision of computing resources.

The consultation was held from January to February 2015. It was targeted at key stakeholders including Principal Investigators and users of current computing projects either funded or hosted by STFC; Principal Investigators of projects likely to have major future computing requirements; Hartree Centre users/partners; CCP Chairs; relevant STFC staff; and representatives of organisations that collaborate with STFC on computing (or may do so in the future). The PPAN advisory panels and Diamond/ISIS facility user committees were asked to provide input on behalf of their communities (the CLF user community input was provided separately). 64 responses (including the advisory panels/user committees) were received. Analysis of the consultation responses informed the Review programme area sub-sections on current plans and long-term requirements; issues and opportunities; and the section on cross-programme topics.

Document review

The Computing Strategic Review Working Group reviewed relevant documents (strategies, reviews, policy documents etc.) from both within STFC and externally (other Research Councils, RCUK, BIS, other relevant national and international organisations). This was used to provide context for the Strategic Review and to ensure alignment with national and international priorities.

The draft Strategic Review was reviewed by the STFC Computing Advisory Panel and the National Laboratory Directors and revisions incorporated before the Review was finalised.

Annex 2: Glossary

ARCHER	UK national supercomputing service
BBSRC	Biotechnology and Biological Sciences Research Council
BIS	Department for Business, Innovation and Skills
CAD	Computer Aided Design
CAE	Computer Aided Engineering
CAM	Computer Aided Manufacturing
CCFE	Culham Centre for Fusion Energy
CEMS	Climate and Environmental Monitoring from Space Facility
CERN	European particle physics laboratory
CLF	Central Laser Facility
CCP	Collaborative Computational Project
Diamond	Diamond Light Source
DiRAC	Distributed Research utilising Advanced Computing (HPC facility)
EC	European Commission
Emerald	GPU supercomputer
EPSRC	Engineering and Physical Sciences Research Council
ESO	European Southern Observatory
GPU	Graphics Processor Unit
GridPP	UK computing grid for particle physics
HPC	High Performance Computing
HTC	High Throughput Computing
ISIS	neutron and muon source facility
JASMIN	super-data-cluster
LHC	Large Hadron Collider
LSST	Large Synoptic Survey Telescope
MSC	Microelectronics Support Centre
MRC	Medical Research Council
NERC	Natural Environment Research Council
NSCCS	National Service for Computational Chemistry Software
PPAN	particle physics, astronomy and nuclear physics
PRACE	Partnership for Advanced Computing in Europe
RAL	Rutherford Appleton Laboratory
R&D	Research and Development
SCARF	Scientific Computing Application Resource for Facilities (STFC compute cluster)
SCD	Scientific Computing Department
SKA	Square Kilometre Array (telescope)
STFC	Science and Technology Facilities Council

Annex 3: Summary of Responses to the Survey of current computing facilities and services

Part 1 - Facilities, Scientific Computing Department and the Hartree Centre

Tables of computing facilities and services provided by the Scientific Computing Department, the Facilities and the Hartree Centre:

Table 1 - HPC/HPC Support

Table 2 – Data Services

Table 3 – Data Processing/Visualisation/Analysis

Table 4 – Software Development, Training and Community Support Activities

Table 5 – Networks

Table 6 – Facility Operations

Note a particular facility or service may be listed in more than one table.

The tables may not be complete.

Table 1 - HPC/HPC Support and HTC

Description	What	Purpose	Location	Date commissioned to end of current funding phase	Users	User numbers
Supporting STFC facilities or science communities:						
APSV2	HPC cluster	HPC support for ASTeC core programmes and some ASTeC funded academic work	Daresbury	2013 -	STFC/ Academic	~30
CLF Computing	HPC servers (and other computing support for CLF – see below)	Support CLF facility operations and experiments	RAL	Commission dates from 2000 – 2013, most systems have no fixed end date	Academic, CLF staff	From around 6 up to around 200/year for each system

Description	What	Purpose	Location	Date commissioned to end of current funding phase	Users	User numbers
SCARF	HPC cluster	Provide HPC resource for STFC staff and collaborators and users of STFC facilities, e.g. for molecular simulation work by ISIS or SCD staff	RAL	2008 – March 2015	Mainly academic, one small group of commercial users	357 registered, 128 active in last year
Supporting research communities funded by other Research Councils:						
Emerald	HPC cluster	Provides a large-scale GPU resource and supports the porting of codes to GPUs to enable production work using GPUs and enable existing codes to use GPUs for improved power efficiency	RAL	2012 – March 2015	Academic, SE5 consortium users	273 registered users
JASMIN	Data intensive computing platform (HPC, data storage, virtualisation environment)	Provides a high-end computing platform for the environmental science community including data management and data analysis	Central hub at RAL with small satellite centres at Reading and Leeds	2012 – April 2017	Academic, public sector, some industry	30,000+ data centre users worldwide; 400+ direct login users
National Service for Computational Chemistry Software (NSCCS)	HPC resource	STFC provides a managed HPC resource for the service to run computational chemistry codes; applications and user support provided by NSCCS at Imperial College	RAL	2011 – 1 Feb 2016	Academic	Varies, currently 241 active users

Description	What	Purpose	Location	Date commissioned to end of current funding phase	Users	User numbers
Supporting national services/cross-Council communities:						
ARCHER System Support	Support for national HPC service run by the ARCHER partners (see other funders)	ARCHER supports researchers run simulations and calculations that require HPC resource	Edinburgh	2013 -	Academic	n/k
GridPP Tier 1 (NB. Also included in Programmes summary tables)	HTC	Supports exploitation of LHC and other particle physics experiments	RAL	2001 – March 2015	Academic	Not known – integral part of the Worldwide LHC Computing Grid so difficult to estimate global users
Hartree Centre systems	HPC systems	Modelling, simulation, visualisation, bigdata analytics and energy efficient computing	Daresbury	2012 – March 2020	Industry and academic	1000 registered

Table 2 - Data Services (accounting/configuration data management, data storage, back-up and disaster recovery services, databases and repositories, data management)

Description	What	Purpose	Location	Date commissioned to end of current funding phase	Users	User numbers
Supporting STFC facilities or science communities:						
APEL	Accounting system	Collects accounting data from sites into database; generates statistical summaries. Supports LHC and EGI communities.	RAL	2013 -	Academic	n/k (data received from 6555 users; data collected from 288 Grid/HPC sites in EU; LHC jobs run on behalf of ~5000 LHC physicists
CLF Computing	File stores; HPC servers; laser control systems; instrument, safety and remote positioning systems; data management and analysis services	Support CLF facility operations and experiments	RAL	Commission dates from 2000 – 2013, most systems have no fixed end date	Academic, CLF staff	From around 6 up to around 200/year for each system
CLF Data Management Service	Data service	Catalogues data and metadata from the Astra-Gemini laser and enables it to be displayed in near real-time	RAL	2011 – April 2015	Primarily academic	~10 CLF operators and ~50 facility users p.a.
Database Services for ISIS	Data service (24 Oracle databases)	Supports ISIS activities (proposal management, user / account management, file metadata, content management system backend)	RAL	n/k – March 2015	Academic	n/k

Description	What	Purpose	Location	Date commissioned to end of current funding phase	Users	User numbers
ISIS Back-up Services (DMF, SDB)	Data storage software	Data Migration Facility – migrates data from disk to tape library; Safety Deposit Box – for the digital preservation of data (extracts and stores metadata on user's files)	RAL	DMF: 2006 – Dec 2017; SDB: 2011 -	Academic (and STFC for SDB)	n/k, 54 registered groups for DMF, 1 group (ISIS) for SDB
ISIS Data Archive Service	Data archive	Long-term data curation service for ISIS as critical component of their disaster recovery planning	RAL	2010 – April 2014	Academic	n/k
Data Services for MSSL (Mullard Space Science Laboratory)	Data service	Ingests solar data from KEK in Japan	RAL	2007 – March 2015	Academic	n/k – may be on behalf of hundreds of users
ePubs	Archive of research publications produced by STFC staff or produced using STFC RAL facilities	Provides external access to scientific and technical output of STFC and management information on publications produced by STFC.	RAL (hardware), service open to all	2012 – (annual funding agreements)	n/k but probably a mixture of academic/ industry/ other public sector	Over 14,000 in previous 12 months
Facilities Data Service for DLS	Data service	Off-site access for end users of the DLS data archive; disaster recovery support for Diamond if other copies of data lost	RAL	2011 -	Academic	~5000 total
Grid Operations Configuration Database (GOCDB)	Configuratio-n management database	Registers GridPP and EGI.eu configuration data (member sites and services, users, service down-times contact details etc.)	Website developed at RAL	2014 – 1 May 2016	Academic	2457 registered

Description	What	Purpose	Location	Date commissioned to end of current funding phase	Users	User numbers
UK Cluster Data Centre	Data processing and archive system	Produces and disseminates quicklook and survey data for the ESA Cluster mission. Also provides contextual information from other missions.	RAL	2000 – end 2016	Academic	n/k
UK Solar System Data Centre	Data service	Provides data archives for the UK solar system community; also data processing and database services for various solar satellite missions	RAL	2005 – March 2016	Academic, industry, general public	~8200 registered
Supporting research communities funded by other Research Councils:						
BBSRC Data Service	Data Service	Data back-up service for BBSRC institutes	RAL	2005 – March 2015	Academic	n/k
CEDA Data Service	Data service	Stores data for CEDA	RAL	2011 – March 2015	Primarily academic; possibly some other public sector and industry	n/k
Chemical Database Service	Data service	Provides chemistry-related databases and associated software to UK academic community. Major component CrystalWorks.	Daresbury	2013 – August 2015	Academic, Other public sector (research council, charities' research facilities)	~6000 registered

Description	What	Purpose	Location	Date commissioned to end of current funding phase	Users	User numbers
JASMIN	Data intensive computing platform (HPC, data storage, virtualisation environment)	Provides a high-end computing platform for the environmental science community including data management and data analysis	Central hub at RAL with small satellite centres at Reading and Leeds	2012 – April 2017	Academic, public sector, some industry	30,000+ data centre users worldwide; 400+ direct login users
MRC Gateway to Research	Data service	Search portal enabling users to find metadata about MRC's cohort studies plus richer, variable level information when logging in	Service hosted at RAL	2009 – March 2016	Academic	n/k
Supporting national services/cross-Council communities:						
Data Service for RCaH (Research Complex at Harwell)	Data service	Provides data storage/backup for Research Complex at Harwell	RAL	2012 – March 2015	Primarily academic, also industry	n/k

Table 3 - Data Processing/Visualisation/Analysis

Description	What	Purpose	Location	Date commissioned to end of current funding phase	Users	User numbers
Supporting STFC facilities or science communities:						
Atlas Visualisation Facility	3D visualisation facility	Data visualisation and analysis	RAL	2013 – April 2015	STFC (Facilities, SCD, Hartree)	n/k (over 500 people in first 18 months)
UK Cluster Data Centre	Data processing and archive system	Produces and disseminates quicklook and survey data for the ESA Cluster mission. Also provides contextual information from other missions.	RAL	2000 – end 2016	Academic	n/k
UK Solar System Data Centre	Data service	Provides data archives for the UK solar system community; also data processing and database services for various solar satellite missions	RAL	2005 – March 2016	Academic, industry, general public	~8200 registered
Supporting research communities funded by other Research Councils:						
GERB Ground Segment Processing System (GERB = Geostationary Earth Radiation Budget)	Data processing system (for the GERB instrument on the Meteosat Second Generation weather satellite)	Processes raw GERB data and makes data available to the GERB science team and wider scientific community	RAL	2002 – August 2016	Academic	n/k
Supporting national services/cross-Council communities:						
Hartree Centre systems	HPC systems	Modelling, simulation, visualisation, bigdata analytics and energy efficient computing	Daresbury	2012 – March 2020	Industry and academic	1000 registered

Table 4 - Software Development, Training and Community Support Activities

Description	What	Purpose	Location	Date commissioned to end of current funding phase	Users	User numbers
Supporting STFC facilities or science communities:						
UK eScience Certification Authority	Digital certification authority	Issues X509 digital certificates to the UK academic community for eScience, Grid and Cloud activities.	RAL (and some staff at Daresbury)	2009 – March 2015	Academic	Certificates valid for one year. ~2400 current valid certificates.
Supporting research communities funded by other Research Councils:						
CCP Support	Community scientific and admin support	Supports computational science activities through software development, co-ordination, training and other activities	Staff based at RAL, Daresbury	1980s – April 2016	Primarily academic; some industry	n/k – probably reach a few thousand UK people and significant numbers overseas
National Service for Computational Chemistry Software (NSCCS)	HPC resource	STFC provides a managed HPC resource for the service to run computational chemistry codes; applications and user support provided by NSCCS at Imperial College	RAL	2011 – 1 Feb 2016	Academic	Varies, currently 241 active users
Software Engineering Support Centre (SESC)	Education, training, support and facilities related to software engineering to EPSRC community engaged in software development	Improve quality and sustainability of software developed within UK computational projects.	RAL	2013 – July 2018	Mainly academic	Figures for CCPForge Service (one element of the SESC): 3070 registered users, 108 registered projects
Supporting national services/cross-Council communities:						
Hartree Centre systems	HPC systems	Modelling, simulation, visualisation, bigdata analytics and energy efficient computing	Daresbury	2012 – March 2020	Industry and academic	1000 registered

Table 5 - Networks

Description	What	Purpose	Location	Date commissioned to end of current funding phase	Users	User numbers
Supporting STFC facilities or science communities:						
STFC Network Service	Computer networking across STFC	Underpin the successful operation of STFC Programmes	Across STFC	Year-on-year operation	STFC staff, facility users, tenants and visitors	n/k

Table 6 - Facility Operations

Description	What	Purpose	Location	Date commissioned to end of current funding phase	Users	User numbers
Supporting STFC facilities or science communities:						
CLF Computing	HPC servers (and other computing support for CLF – see below)	Support CLF facility operations and experiments	RAL	Commission dates from 2000 – 2013, most systems have no fixed end date	Academic, CLF staff	From around 6 up to around 200/year for each system
Diamond Light Source Computing	HTC/HPC. Intel CPUs, NVidia GPUs, parallel storage	Support for scientific feedback on user experiments on Diamond beamlines. Acceleration development simulations. Limited post-visit data analysis.	RAL	2007-	Academic, industry	~5000

Part 2 – Particle Physics, Astronomy and Nuclear Physics Research Programme

HPC Summary

HPC support for Programmes comprises one national facility (DiRAC) covering all PPAN science (mainly astronomy, cosmology and particle physics; a few nuclear physics users) and various HPC clusters/shares of university HPC facilities supporting astronomy research. The main activities the facilities are used for are modelling, simulation and data reduction and analysis. Most systems were commissioned between 2008 and 2012. The end dates for current funding phases are around 2016 – 2018, but in several cases there is no specific end date (e.g. if the system was given capital funding by STFC to set it up but is now maintained by the university).

The users are almost entirely academic. DiRAC supports a small percentage of non-academic users, up to a maximum of 8%. It is difficult to gauge the number of users. There are around 1200 registered users in total, but this will include duplicates and not all will be active users. Where numbers of registered and active users are given, only about a half of the registered users seem to be active.

The total capital funding from STFC for these HPC facilities was £1.7M. Only four facilities receive resource funding from STFC – DiRAC (by far the largest at £1.3M p.a.), Queen's University Belfast, UCL and the University of Sheffield HPC cluster. Several of the university clusters are funded by the university. Other funding sources include BIS, the EU, ERC, UKSA and the Royal Society.

Apart from the primary purpose of supporting PPAN research, there were few additional activities supported by the services. The few examples included training (DiRAC, Sussex, UCL) and code development (St Andrews).

HPC - Some STFC funding

Name	What	Purpose	Location	Date commissioned to end of current funding phase	Users	User numbers
DiRAC	HPC facility	Support PPAN research in theoretical modelling, data modelling and numerical simulation	Cambridge, Durham, Edinburgh, Leicester	2012 – March 2017	Mostly academic (up to 8% available to non-academics)	630 registered, 260p.a.
Belfast HPC Cluster	HPC cluster	Data reduction, analysis and modelling for astronomers (solar, exoplanet, supernovae)	Belfast	2008 – March 2017	Academic	60-70
UH STRI Cluster	HPC cluster	Supporting astrophysics observational and modelling work	Hatfield, Hertfordshire	2009 – April 2018	Academic	134 registered (60 active?)
ALMA Cluster	HPC cluster	Data reduction from the ESO ALMA telescope	Manchester	2012 -	Academic	7 local plus all UK ALMA users
COMA Cluster	HPC cluster	Cosmology and extra galactic modelling and simulation work	Manchester	2009 -	Academic	43 registered
HYDRUS/ ADDER/ JUMPER	HPC clusters	Real-time observations of pulsars using Lovell Telescope and offline searching for pulsars from surveys	Manchester	2008 -	Academic	29
QMUL Cluster	HPC cluster	Support simulation of planetary formation, space plasma and other research activities	London (Queen Mary)	2014 – April 2016	Academic	~10p.a.
Iceberg Cluster	Share of HPC facility	University-wide HPC facility, used by Sheffield astronomers (some parts reserved for STFC projects)	Sheffield	2011 – Sept 2016	Academic	n/k (950 registered in total)

Name	What	Purpose	Location	Date commissioned to end of current funding phase	Users	User numbers
Apollo HPC Cluster	Share of HPC cluster	Observational data reduction, simulation data reduction and theoretical modelling for astronomers	Brighton, Sussex	2010 -	Academic	48 registered
UCL HPC Cluster	HPC clusters (4 clusters)	Supports astrophysics research including simulations, data analysis and processing	London (UCL)	1. 2010 – 2012; 2. 2011 – April 2017; 3. 2010 – Dec 2017 4. 2015 – Dec 2019	Academic	1. 95+ have access; 2. ~80; 3. ~20 4. ~15

HPC - No STFC funding

Name	What	Purpose	Location	Date commissioned to end of current funding phase	Users	User numbers
e-MERLIN Correlator	Telescope correlator	Combines signals from the e-MERLIN array of radio telescopes into data products	Manchester	2010/2012 – 2017	Academic	N/A (provides data processing for 500+ users)
Fornax	HPC cluster	Supports cosmology and extra galactic modelling and simulation work	Manchester	2014 -	Academic	25 registered
Glamdring Cluster	HPC cluster	Supporting astronomy research programmes, many involving intense data reduction	Oxford	2010 -	Academic	110 registered
St Andrews DiRAC 1 Cluster	HPC cluster	HPC support for UK MHD community	St Andrews	2009 – Nov 2012	Academic	85 registered (50 active)

HTC Summary

HTC support comprises GridPP plus some gravitational wave support at Cardiff. GridPP primarily supports the LHC experiments but also supports T2K and around 30 Virtual Organisations (VOs) (e.g. other particle physics projects such as ILC, IceCube, SNO+; the theoretical community (Phenogrid); and non-particle physics research (BioMed)). GridPP Tier 1 is based at RAL with Tier 2 distributed across 19 UK sites, the largest at Glasgow, Imperial, Lancaster, Manchester, Queen Mary and RAL. The capital funding for the current phase (GridPP4 – ends March 2015) was around £6.8M, with resource funding for staff and operations of around £5M p.a.. There are also significant (in some cases major) contributions from many of the university host sites. The university contributions cover costs such as providing the machine rooms, operating costs, hardware and networking. STFC resource contributions fund Tier 2 operations staff, experiment support posts, travel and consumables. Funding has been approved for GridPP4+, covering April 2015 to March 2016.

The vast majority of GridPP users are academic. The number of users cannot be measured as much scheduled production work is done 'on behalf of' many thousands of 'users' using what are effectively production accounts; and many thousands of users also submit individual jobs to the worldwide LHC computing Grid of which GridPP is a part. If all LHC publications use LHC computing resources, then all LHC authors are either directly or indirectly users. The LHC experiments amount to around 7000 authors of which the UK represents about 10%, giving the figure of 700 users. But the non-LHC VOs also have significant numbers of authors.

The gravitational wave computing at Cardiff supports 20 active users at a cost of £300k p.a. with the current phase of funding due to end in December 2015.

Name	What	Purpose	Location	Date commissioned to end of current funding phase	Users	User numbers
GridPP Tier 1	HTC	Supports exploitation of LHC and other particle physics experiments	RAL	2001 – March 2015	Academic	Not known – integral part of the Worldwide LHC Computing Grid so difficult to estimate global users
GridPP Tier 2	HTC	Supports exploitation of LHC and other particle physics experiments	Distributed (19 UK sites)	2001 – March 2015	Almost all academic, very minor industry/ health involvement	Not known (see above) – ~700 UK users based on estimate of UK LHC author numbers
Gravitational waves computing	HTC (share of Cardiff HPC/HTC cluster)	Computing for UK gravitational wave science	Cardiff	2013 – December 2015	Academic	20 active, 50 total

Astronomy Survey Data Centres

The XMM Survey Science Centre processing system is due to close by the end of 2015. The CASU Data Processing System provides expert pipeline processing for key ground and space-based imaging and spectroscopic surveys. The Wide-Field Astronomy Unit archives and publishes sky survey datasets and also generates additional advanced data products from the files produced by CASU data reduction pipelines. STFC provides £1M p.a. resource funding for CASU and WFAU, which are also supported by UKSA. The users are currently all academic. There are over 1000 registered users of WFAU. The number of users of CASU is not known but includes all users of ESO public survey data, all VISTA users, all WFCAM users, and all users of Gaia etc. Both units are likely to be involved in the delivery of data services for LSST/Euclid and SKA.

Name	**What**	**Purpose**	**Location**	**Date commissioned to end of current funding phase**	**Users**	**User numbers**
CASU Data Processing System	Data processing facility	Pipeline processing for key ground-based imaging and spectroscopic surveys	Cambridge	1998 – March 2016	Academic	n/k
Wide-Field Astronomy Unit	Data management service	Archives optical and near-infrared sky survey data on behalf of the UK astronomical community and that of ESO	Edinburgh	1999 – March 2016	Academic, possible future industry collaboration	>1000 registered users for proprietary data access to WFCAM Science Archive; more users for public WFAU data
XMM Survey Science Centre processing system and website	Data processing cluster and website	Generates daily XMM pipeline science products and catalogues for delivery to ESA and the public	Leicester	2007 – April 2011 (end of STFC funding – system due to close by end 2015)	Academic	~10