



SCIENTIFIC COMPUTING DEPARTMENT STRATEGY 2017 – 2021



Science & Technology
Facilities Council

Our Mission

The mission of the Scientific Computing Department (SCD) is:

“to maximize the impact of scientific computing through our expertise, leadership and collaboration.”

Our mission is fully aligned to support and enable the STFC Corporate Strategy and the STFC e-infrastructure Strategy.

The Scientific Computing Department fulfils its mission by:

- Designing, deploying and operating large and complex computing and data systems
- Supporting the research life-cycle by extracting insights and value from data
- Creating algorithms and software to exploit future research computing infrastructure
- Providing cross-domain expertise to develop, innovate, and sustain software, and related digital assets for research
- Leading and participating in the national and international collaborations working to achieve these aims
- Committing to STFC's corporate theme of Inspiring and Involving to deliver a vibrant programme of public engagement
- Actively supporting and contributing to the developing skills agenda in Scientific Computing

For further information and discussion please contact the Scientific Computing Department Director: Tom Griffin (tom.griffin@stfc.ac.uk)

Delivering our mission

SCD is one of the largest scientific computing departments in Europe and comprises around one hundred and fifty staff. We deliver expertise in computational science and professional large-scale scientific and data management and computing systems, services and expertise to our STFC, national, and international scientific user communities, and to our collaborators and stakeholders in other Research Councils – particularly the Engineering and Physical Sciences Research Council (EPSRC), the Natural Environment Research Council (NERC), the Biotechnology and Biological Sciences Research Council (BBSRC), and the Medical Research Council (MRC).

Delivering our mission builds on our core strengths and is described in the following sections.

- 1. SCD will build and strengthen our existing capability of providing enabling computational expertise and e-infrastructure to support STFC in fulfilling its strategic goals: “world class research, world class innovation, and world class skills.”

When busy processing LHC data, the Tier-1 moves about 1 Petabyte a day between compute and storage system.

- 1.1. The GridPP collaboration provides the UK component of the Worldwide Large Hadron Collider Computing Grid. SCD delivers the major part of the UK share of the distributed computing infrastructure required for analysis of data from the Large Hadron Collider by hosting one of the twelve ‘Tier 1’ international centres that provides custodial data storage and supports analysis of data from the Large Hadron Collider experiments based at CERN in Geneva.

SCD’s UK Tier-1 has a 60 petabyte tape and disk storage system and would be able to hold the equivalent of about 18 million HD feature films, probably more than 30 times the number ever produced by the film industry.

- 1.2. In collaboration with the Natural Environment Research Council (NERC), the UK Space Agency, and RALSpace, SCD has designed, built and manages the highly innovative JASMIN Super-Data-Cluster which combines Petabyte storage capability closely coupled to a high performance computing cluster. JASMIN is used by 30,000 users from the world-wide climate modelling community and other environmental disciplines.

JASMIN has international commitments to store and process data on behalf of organisations such as the European Space Agency and NASA. JASMIN is the UK’s primary scientific environmental data analysis platform. It represents a £16M investment and is funded by a wide range of collaborating agencies and institutes : NERC, STFC, UK Space Agency, UK Met Office Leeds, Reading, Oxford, Bristol. Edinburgh Universities, while also enabling UK industry access to environmental monitor data.

- 1.3. Through funding from the Engineering and Physical Sciences Research Council (EPSRC), the Biotechnology and Biological Sciences Research Council (BBSRC), and the Medical Research Council (MRC), SCD plays a vital role in providing expert, long-term scientific software support and stewardship for the UK scientific research user communities engaged in UK’s Collaborative Computational Projects (CCPs), and to users of the UK’s national supercomputing service.

- 1.4. SCD provides computational science and engineering expertise and leadership in mathematics, physics, chemistry, biology and engineering. Funded mainly by EPSRC, and more recently also by the Hartree Centre, and STFC large experimental Facilities, the primary focus of these groups is on the development and maintenance of major computational science codes used by thousands of users in both academia and industry. This includes research and development of algorithms and theoretical computational models with an emphasis on optimizing these codes for execution on modern high performance computing systems.

Over 4000 developers have downloaded HSL mathematical software library in the past 3 years.

- 1.5. SCD is heavily engaged with international partners in Europe, North America and throughout the world. We provide software and services to international collaborations supported by STFC, such as the CERN Large Hadron Collider, the Square Kilometre Array radio telescope, and European facilities such as the European Synchrotron Radiation Facility and the Institut Laue-Langevin. SCD has a long track record in participating in the development of a European research infrastructure, including the Partnership for Advanced Computing in Europe (PRACE), the European Grid Initiative, and EUDAT European research data infrastructure, and by leading the European Open Science Cloud Pilot.

- 1.6. SCD is committed to STFC’s strategic theme “Inspiring and Involving”. We will contribute to this by using our communities and facilities as the basis of public engagement to demonstrate how the work of SCD enables our scientific endeavours, and secondly by directly engaging them with ways that computers can be used to analyse and manipulate data.

The Ada Lovelace Centre will transform research at large facilities through a multidisciplinary approach to data processing, computer simulation and data analytics. The impact will be felt by industry and academia across the many science disciplines these facilities support,

- 2. SCD is committed to deliver and further develop its expertise in scientific analysis, data science, modelling and simulation, to support STFC’s commitment to: *Establish the Ada Lovelace Centre, an integrated, cross disciplinary, data-intensive science centre, to transform the use of real time data processing, computer simulation and data analytics to deliver more effective research at our national facilities.*

- 2.1. Over 10,000 users of the STFC Facilities – Diamond Light Source, ISIS neutron and muon sources and Central Laser Facility – already benefit from the data and compute services SCD delivers. We will build on these core strengths, and further develop our collaboration with the STFC facilities in the

development of analysis and simulation tools to support end users. This will form the foundation for SCD’s contribution to the Ada Lovelace Centre, for which we are working with the RCUK e-infrastructure group to seek funding.

- 2.2. The Ada Lovelace Centre will significantly enhance our capability to support the Facilities’ science programme, with a focus on their needs through:
 - Building capacity in advanced software infrastructure for the handling, analysis, visualisation, integration, modelling, and interpretation of experimental data
 - Building and curating science domain databases of analysed and published data to enable long-term exploitation of the data
 - Developing the skills in data and software experts educated at the frontier of their disciplines in the fast expanding field of data intensive science
 - Supplying computational hardware to enable the data intensive analysis, simulation and modelling to be performed.

- 3. SCD delivers a comprehensive programme of computational and data science services, research and development to underpin STFC’s Data Intensive Science ambition: *to develop and deliver cutting-edge solutions for academia and industry to advance data intensive science and innovation.*

- 3.1. SCD will continue to provide vital science and software expertise to enable the Hartree Centre to fulfil its mission to: “Transform the competitiveness of UK industry by accelerating the adoption of data-centric computing, big data and cognitive technologies”. Collaborating with the Hartree Centre, SCD will also carry out innovative technology development to ensure best use is made of novel and emerging software and hardware technologies in scientific computing across STFC and its collaborators.

SCD are an integral part of industrial projects run by the Hartree centre. Both contract and collaborative research approaches are deployed to support utilisation of existing scientific computing methodologies, and to develop innovative new methods of value to industrial users.

The ICAT suite of data management tools are developed and maintained by SCD. These include tools for data cataloguing, data access and data download. They are used within ISIS, Diamond Light Sources, and Central Laser Facility. ICAT is being developed as an open source collaboration and is being used with within facilities around the world.

3.2. The UK-T0 is collaboration across STFC programmes in Astronomy, Astrophysics, Cosmology, Nuclear Physics, Particle Physics and Solar System science. UK-T0 includes the STFC facilities, and also the Culham Centre for Fusion Energy. The science programmes and facilities activities are all data intensive, requiring very large capacity computing and data handling systems. We will work across STFC, its facilities, (and with the Ada Lovelace Centre), and programmes, and with the Culham Centre for Fusion Energy to develop a coordinated, efficient and cost effective e-infrastructure to meet the needs of UK-T0 participants where it makes sense to do so. For STFC this will be a major step forward for the coordination of computing expertise and infrastructure across the entire STFC portfolio - both STFC programmes and facilities. This will create an effective STFC platform for closer engagement with the developing UK e-infrastructure.

- 3.3. SCD will continue to support and innovate in enterprise scale scientific data management. SCD has, over many years, developed a reputation in supporting dependable and scalable data management, within the Tier 1 centre for GridPP, and in supporting the Facilities programme. We will continue to support these tools, champion their use both within STFC and internationally, and to evolve them into the next generation of data management infrastructure, capable of supporting emerging scientific instruments at scale, and across the scientific lifecycle.
- 3.4. SCD will continue to support the dissemination of scientific outputs for STFC, including literature, data and software. It will engage with RCUK (and the evolving UK Research and Innovation) and internationally, to develop and implement a policy framework for Open Access and Open Science that benefits STFC's science community. The Library Service in conjunction with repository services within SCD will continue to deliver services to underpin this objective, including the continued development of STFCs open access repository "ePubs", and building an "eData" repository for data publication. SCD will further explore publishing the interconnection of research outputs to publish a more complete view of science. (See also 4.2)

The work of STFC staff and the users of its scientific facilities is reported in over 1000 scientific publications each year, across a broad range of subjects.



- 3.5. Development of Research Networking for STFC is essential in order to:
- Support the growth and development of existing and future data intensive computing services within STFC.
 - Implement further the ideas behind fast end to end data transfer networks, already deployed by GridPP Tier-1, JASMIN and DLS in order to enable the research community to better exploit national and international e-Infrastructure.

We will seek funding to deploy pilot networks for key user communities which will demonstrate the capability and ease of access to remote compute that this will enable.

SCD facilitates the advancement of computational science and engineering by providing expertise, software development, and training to UK research communities funded by a number of Research Councils.

4. In order to ensure *that the UK's e-infrastructure supports the country's leading international research status and delivers the data capabilities essential for academia, industry and STFC's science program*, SCD will collaborate with our Research Council UK partners and stakeholders, and with Innovate UK, and JISC, as part of the emerging UK Research and Innovation (UKRI), to shape and build an efficient and effective UK e-infrastructure for UK science and industry.
- 4.1. SCD will build on the emerging UK-T0 high throughput computing collaborations, its software expertise with the Computational Science Centre for Research Communities (CoSeC), and its data analytics expertise with Hartree, JASMIN and the Ada Lovelace Centre in order to contribute to the planning and development of the UK e-infrastructure.

- 4.2. SCD will develop its national and international influence in defining and implementing UK, EU and Global policy towards open science. Building on G7, European and national policy development in which SCD played a leading role in formulating, SCD will implement an open-data science environment for the National Laboratories that puts STFC at the forefront of the global move towards open science. This will be supported through leadership in Research Data Alliance, the European Open Science Cloud and the European Data Infrastructure. (See also 3.4)
- 4.3. SCD will continue to play a leading role in developing and sustaining international e-Infrastructure connected to the UK's infrastructure to ensure that STFC's researchers can maintain their leading role in world science programmes. This will include participating in critical European e-infrastructure collaborations, that provide the fundamental Networking, Compute and Data services for European research. SCD will also explore new and synergistic opportunities and collaborations with newly developing Biological and Environmental science research infrastructures.
- 4.4. SCD will strengthen partnerships with National Laboratories in the USA, such as Lawrence Berkeley National Laboratory, Oak Ridge National Laboratory, Argonne National Laboratory, and Brookhaven National Laboratory in order to share best practise and better support the UK scientific community. To seek and form partnerships with institutes elsewhere in the world, such as China, India, South Africa and Australia, including via programmes such as the Global Challenge Research Fund.

SCD has delivered over 3000 training days in computational science and engineering in 2016/17.



Science and Technology Facilities Council

Rutherford Appleton Laboratory, Harwell, Oxford, Didcot, Oxon OX11 0QX, UK.

T: 44 (0)1235 445000

Sci-Tech Daresbury, Daresbury, Warrington, Cheshire WA4 4AD, UK.

T: +44 (0)1925 603000

Establishments at Boulby Underground Science Facility, Cleveland; Chilbolton Observatory, Hampshire; Daresbury Laboratory, Cheshire; Polaris House (STFC headquarters), Swindon; Rutherford Appleton Laboratory, Oxfordshire; UK Astronomy Technology Centre, Edinburgh.



**Science & Technology
Facilities Council**