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.

Data Analysis as a Service
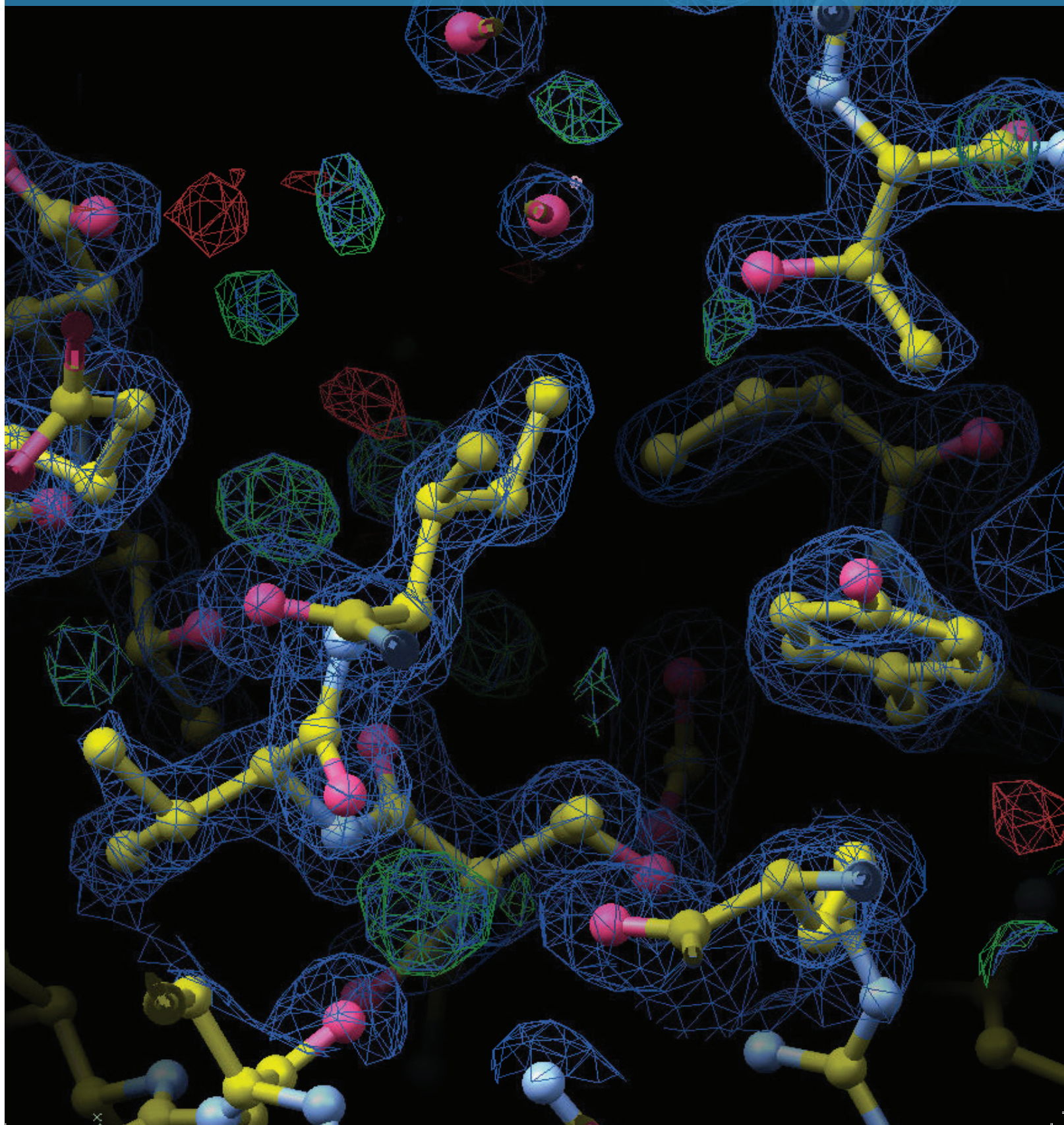
Software underpins the research process: the Software Sustainability Institute 2014 survey [1] of Russell Group [2] Universities showed that 92% of the respondents use research software and 56% of those write their own software. Providing services to enable researchers to use research software and tools to enable better quality software to be written are part of the remit of the Software Engineering Group. The group also aims to reduce the effort required by the facilities in supporting scientists for the analysis process and help increase the scientific output of the facilities.

Data Analysis as a Service

Analysing the scientific data produced by the STFC facilities is becoming more and more complex due to the advancements of the instruments and the scientific techniques. In some cases the volumes of data have grown so large that it is no longer practical for scientists to transport their data back to their home institution. In other cases, the analysis requires access to high performance computing and complex chains of software where these resources and access to the necessary expertise may not be available. All of this technical complexity is being exposed to the scientists, who are quite often not computing experts, and is resulting in the analysis process becoming a bottleneck.

The Scientific Computing Department is working with the facilities based at the Rutherford Appleton Laboratory (ISIS Neutron and Muon Source, the Central Laser Facility and the Diamond Light Source) to implement and deploy a 'Data Analysis as a Service' system (DAaaS) to support scientists during the analysis process. This service provides scientists with easy to use access to compute resources, collocated with the experimental data archives, to efficiently and easily process their data, within a managed, secure virtual environment. Commonly used software packages are systematically made available via a deployment and configuration system, and the environment offered to users is customised according to the nature of the experiment and requirements of the experimental team.

This service aims to reduce the technical complexity of the analysis process by simplifying user workflows and automating complex processes – for instance automating the movement of large volumes of data between computing resources. From this, it will reduce the effort required by the facilities in supporting scientists for the analysis process and help increase the scientific output of the facilities.

One of the leading use cases for DAaaS is the ISIS Data Analysis as a Service project which provides users from multiple communities with all the necessary software, data and compute resources they require to complete their analysis process.

Automated testing of research software

The EPSRC funded Software Engineering Support Centre provides an automated testing facility to the academic community with a focus on the Collaborative Computational Projects. We undertook a survey at the end of 2016 to establish the Continuous Integration practices in the academic community and potential appetite to use a central service. It underpins our planning for the SESC Build Service beyond the current funding and provides an interesting and useful view on the usage of CI and the interest in CI for HPC testing. Full report available from http://purl.org/net/epubs/work/33360356 .

Key points from the survey were:

- 75% of all respondents are already using CI systems
- 81% of those who identified themselves as software developers/RSEs are already using CI systems
- A significant minority of respondents use more than one CI system
- All of the seven UK Research Councils are represented, with EPSRC being the biggest funder
- Python, Fortran, C and C++ are the most common languages reported
- Travis and Jenkins are the most common CI frameworks in use
- Compilers and specific libraries are the most important tools needed for automated testing
- There is an interest in HPC testing, and a smaller interest in novel architectures
- Shibboleth and OpenID are of equal importance as authentication mechanisms

Bringing the two together

One of our aims for the coming year is to enable research software using the automated testing service to be automatically deployed to the DAaaS analysis environments . This will allow developers to release their up to date and well tested software to a large number of users and allow scientists to access the latest software with minimal effort.

Authors
C. M. Jones, F. Barnsley (DAaaS), S. Lamerton (SESC), STFC Rutherford Appleton Laboratory

References
[1] https://www.software.ac.uk/blog/2016-09-12-its-impossible-conduct-research-without-software-say-7-out-10-uk-researchers
    The data collected during this survey is available for download from Zenodo ("S.J. Hettrick et al, UK Research Software Survey 2014", DOI:10.5281/zenodo.14809). It is licensed under a Creative Commons by Attribution licence (attribution to The University of Edinburgh on behalf of the Software Sustainability Institute

[2] The Russell Group comprises the following universities – Birmingham, Bristol, Cambridge, Cardiff, Durham, Edinburgh, Exeter, Glasgow, Imperial College London, King's College London, Leeds, Liverpool, London School of Economics and Political Science, Manchester, Newcastle, Nottingham, Oxford, Queen Mary University of London, Queen's University Belfast, Sheffield, Southampton, University College London, Warwick and York.