Shared Project: The Research Data Transfer Zone

Research institutions are developing ways to optimise data-driven research across distance, connecting researchers to laboratory, compute and storage services via high-speed data networks.

The goal of 'campus bridging' is to seamlessly integrate:
- a researcher's personal research infrastructure,
- the research infrastructure at other campuses,
- the research infrastructure at the regional, national and international levels,

so that they all function as if they were proximate to the scientist achieving efficiency gains and enabling new ways of working.

In the US, the Department of Energy is investing in a *'Science DMZ'* model, whereby high-volume research data is handled in a separate environment to business data. This approach is already established with the GridPP particle physics community. Jisc, though it's Janet End-to-End Performance Initiative is working with research-intensive institutions in the UK to adopt the same principles, using tools that are efficient and easy to use for wider data-driven research community.

The purpose of the shared project was to identify and engage with researchers with significant data transfer requirements and relevant professional services to understand how these needs are being met and evaluate whether a RDTZ approach could form part of an effective data transfer ecosystem. With buy-in from the μ-VIS X-Ray Imaging Centre at the University of Southampton, a demonstrator environment was established by university network services with support from Jisc and DLS. Test data and research data sets from experiments at DLS achieved on average a ten-fold increase in end-to-end performance. The direct and noticeable benefit of this is that researchers maintain their mental momentum, being able to analyse experimental data almost as soon as it is generated, rather than waiting 2-3 weeks.

As the work between Jisc, DLS and SES and its member organisations progress, we expect to see increased pressure on the university connections to Janet as researchers perform higher data rate transfers, and it seems timely for a broad strategic discussion on increasing capabilities to support demand.

This will help to shape a coherent strategy that delivers on the need for an effective and efficient the research e-infrastructure ecosystem, both for the UK across longer distances which are more challenging.
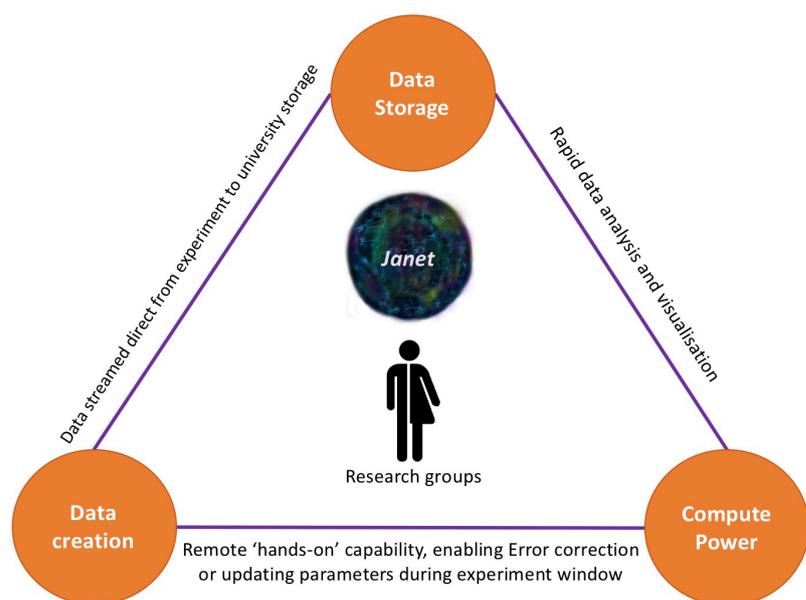

Figure 1: Superfast networks connecting research capabilities

The benefits high-speed data transfer for UK Research

1. Ensure the UK academic community is on a competitive trajectory to conduct world-leading data-intensive research with collaborators in the UK and internationally.
2. Optimise the capability for UK research communities to exploit the growth in capacity of the Janet network as it moves towards Terabit networking.
3. Address the challenge of linking increasingly affordable but high data volume networked scientific equipment, such as electron microscopes and gene sequencers, to national centres for processing and long-term storage of data.
4. Enhance the ability of remote scientists to carry out 'real-time' research activities such as remote experiment control at national and international experimental facilities such as the Diamond Light Source and remote telescopes and observatories such as SKA. Rapid access to data outputs also enable efficiency gains obtained through the ability to identify and resolve errors or make parameter adjustments during the experiment window.
5. Improve the exploitation of the NeI by enabling real-time access to co-located processing capability and data caches such as that exemplified by NERC's JASMIN Super Data Cluster.
6. Provide a capability for universities to build expertise and share best practice in campus network engineering for data-intensive science. This will increase the potential for cross-fertilisation of data science methodologies used by different scientific communities. It will also reach out to disciplines currently unaware of the potential of the Janet network to increase their research output.
7. Provide an open source high-capacity data transfer toolkit that will ensure that users can easily exploit the UK RDTZ infrastructure to its full potential.
8. Make UK researchers internationally competitive in data intensive research applications by provision of world-leading end-to-end research data network performance.
9. Make regional and national research facilities more accessible and hence more attractive to researchers, increasing quality of research, collaborative opportunities.

Further reading

The SES RDTZ event at the University of Southampton

The SES RDTZ webpage, with links to presentation materials, the Jisc End-to-end network Performance Project and the U.S. Department of Energy Science DMZ project

Report Contributors

Timothy Metcalf, Manager for the SES Consortium
Prof. Tony Hey, Chief Data Scientist, Science & Technology Facilities Council
Prof. Simon Cox, CIO, University of Southampton
Dr. Tim Chown, Jisc End-to-End Network Performance Project Manager