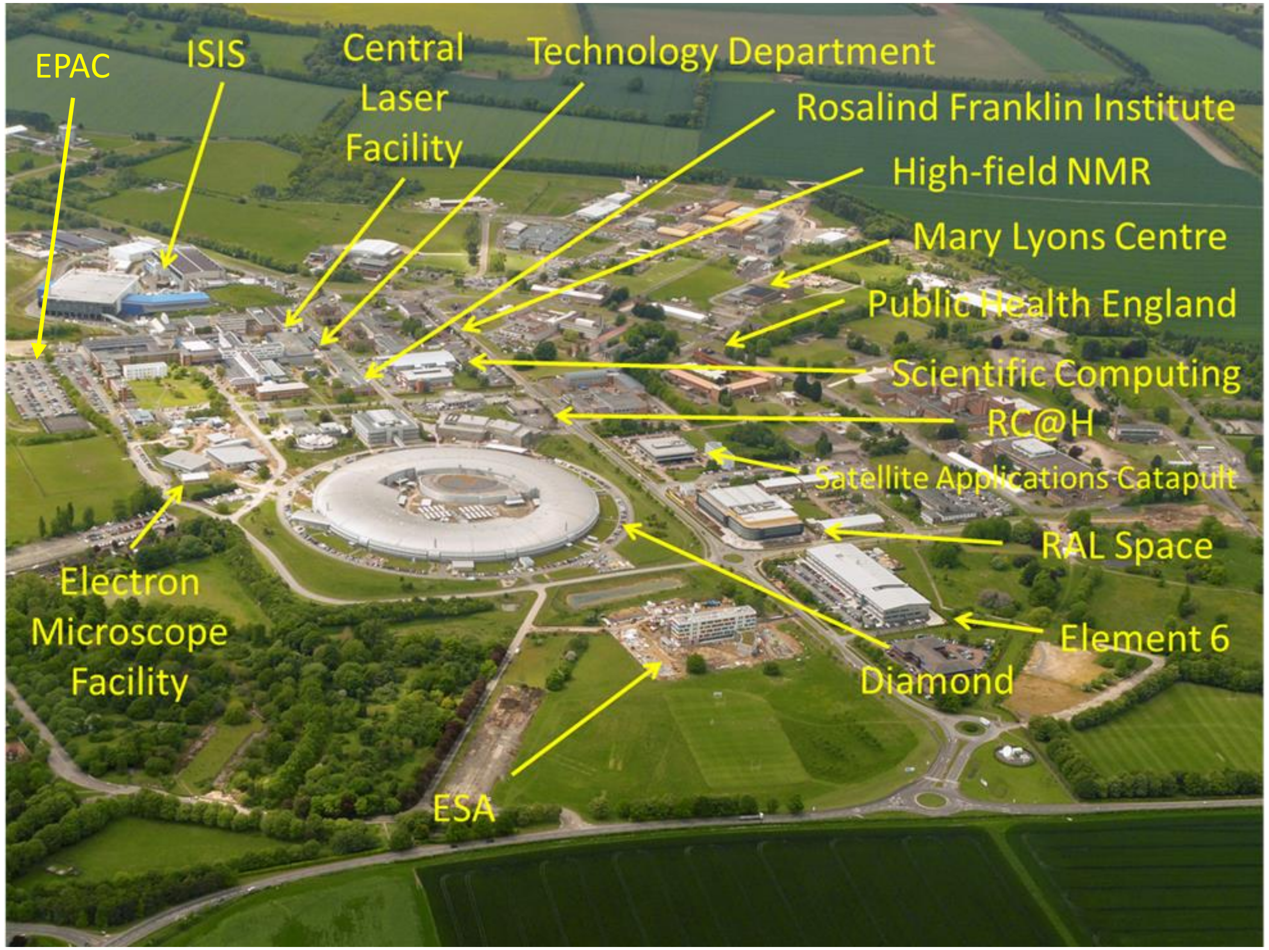
Highly diverse computing (including the long tail)

or

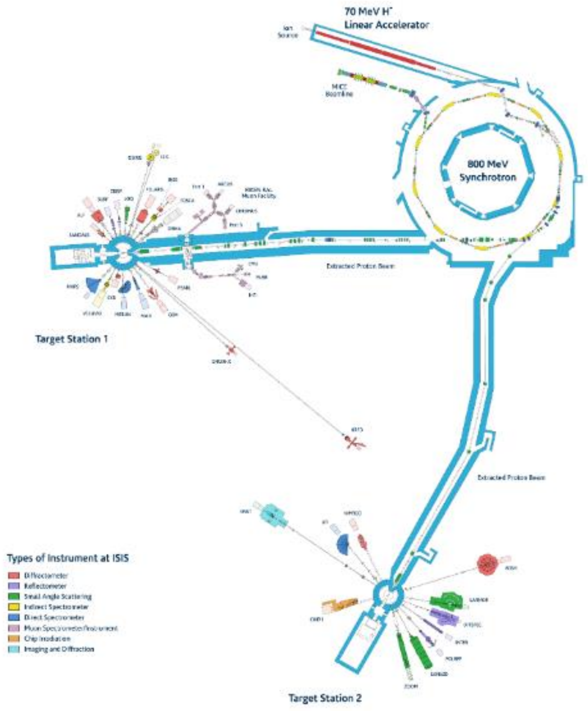
Computing for STFC Facilities

UKRI Science and Technology Facilities Council ISIS

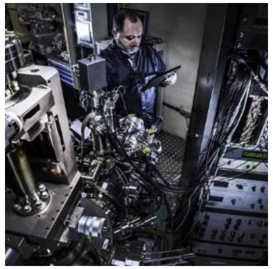
What are the STFC Facilities?

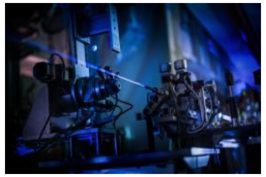


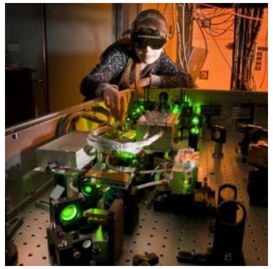
Science and Technology Facilities Council

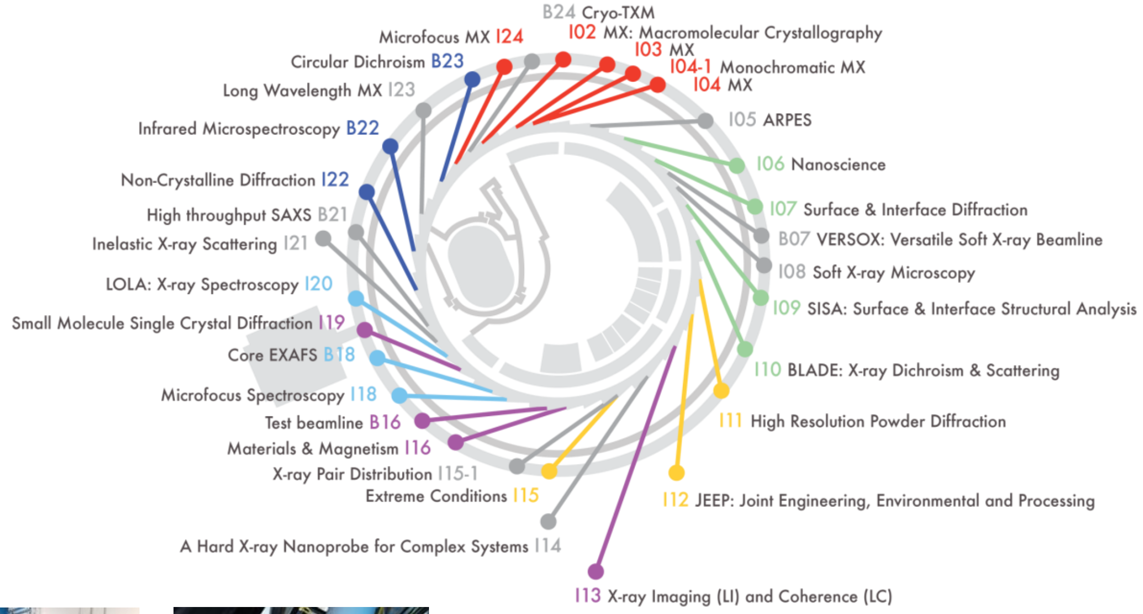


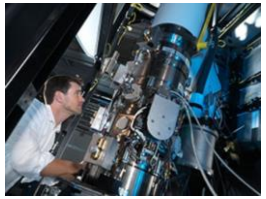
What are the STFC Facilities?








B24 Cryo-TXM
Microfocus MX I24
I02 MX: Macromolecular Crystallography
I03 MX
Circular Dichroism B23
I04-1 Monochromatic MX
I04 MX
Long Wavelength MX I23
I05 ARPES
Infrared Microspectroscopy B22
I06 Nanoscience
Non-Crystalline Diffraction I22
I07 Surface & Interface Diffraction
High throughput SAXS B21
B07 VERSOX: Versatile Soft X-ray Beamline
Inelastic X-ray Scattering I21
I08 Soft X-ray Microscopy
LOLA: X-ray Spectroscopy I20
I09 SISA: Surface & Interface Structural Analysis
Small Molecule Single Crystal Diffraction I19
Core EXAFS B18
I10 BLADE: X-ray Dichroism & Scattering
Microfocus Spectroscopy I18
Test beamline B16
I11 High Resolution Powder Diffraction
Materials & Magnetism I16
X-ray Pair Distribution I15-1
Extreme Conditions I15
I12 JEEP: Joint Engineering, Environmental and Processing
A Hard X-ray Nanoprobe for Complex Systems I14
I13 X-ray Imaging (LI) and Coherence (LC)

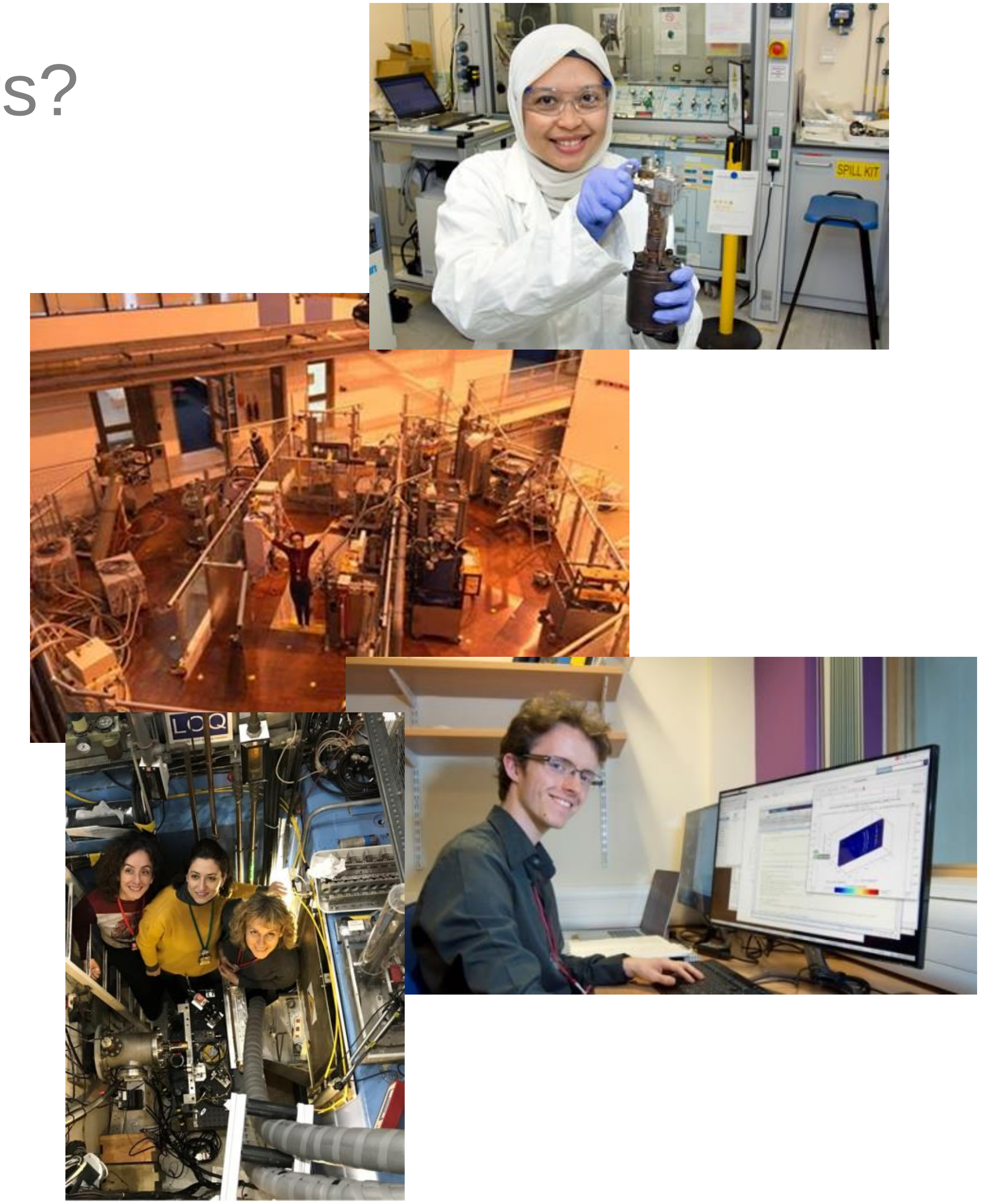


Science and Technology Facilities Council

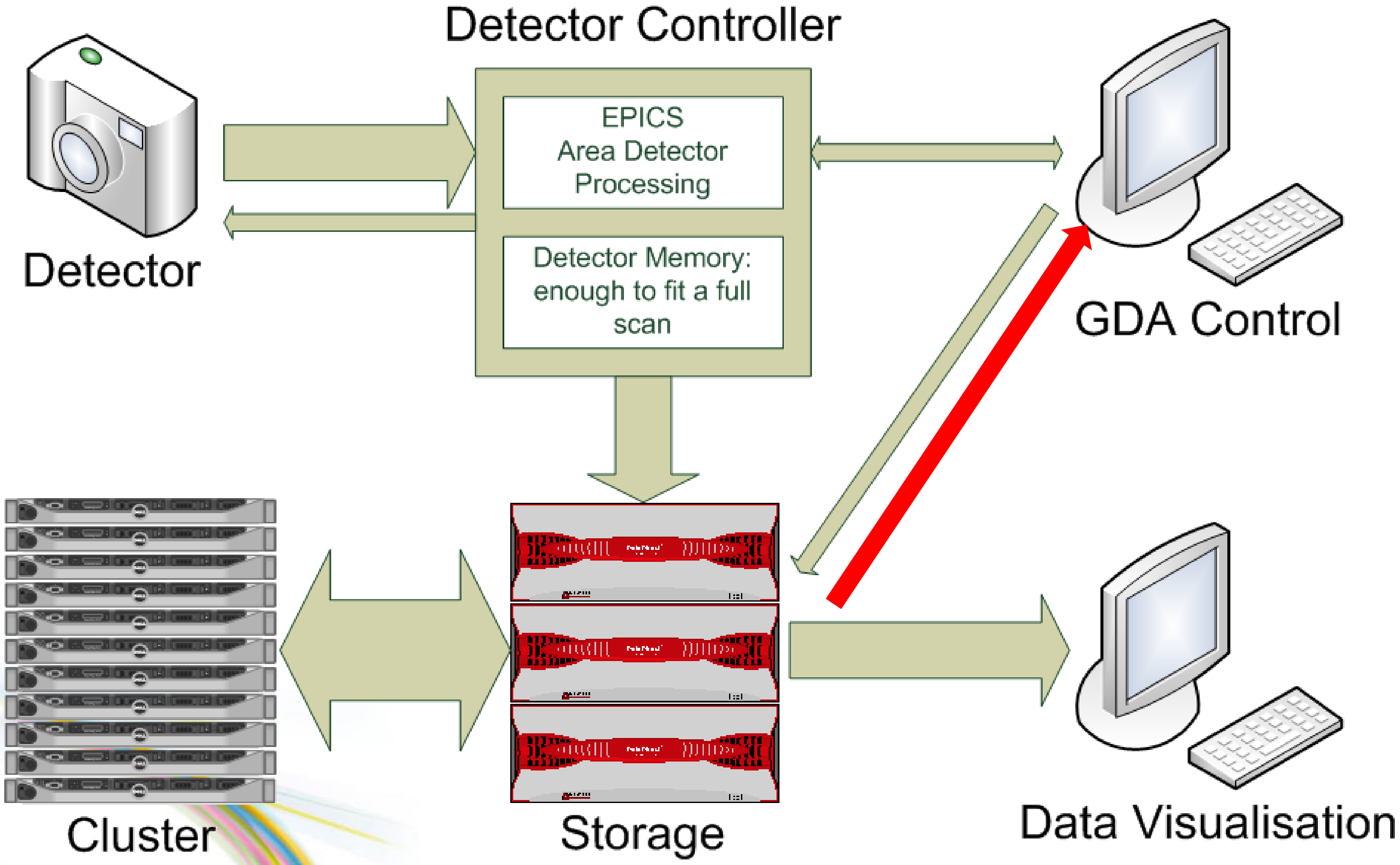
ISIS

Who are the users?

- Diverse …
- Thousands of users – national and international, on-site and remote
- Biologists to engineers
- Experts to non-experts
- 24/7 operation, multiple simultaneous user groups, individual experiments from minutes to weeks
- Continuous development, small and large (eBIC/ePSIC, EPAC, Diamond Lattice upgrade …)

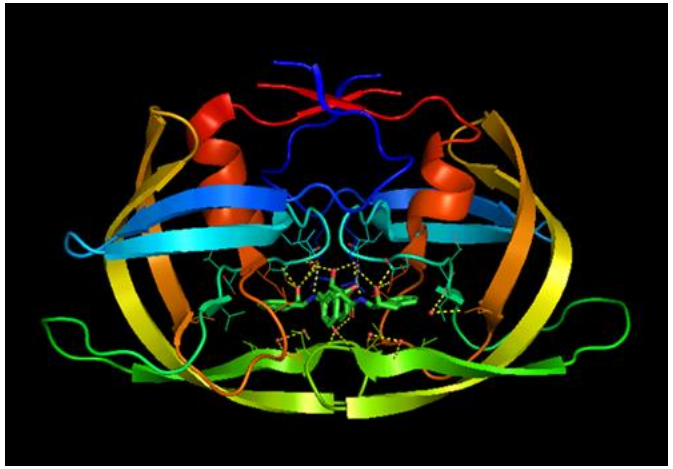


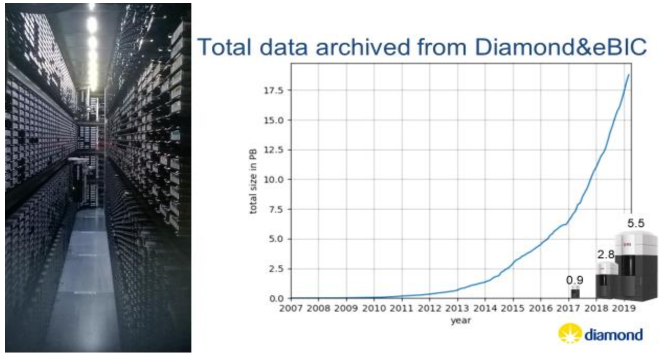
Typical User Data Flow

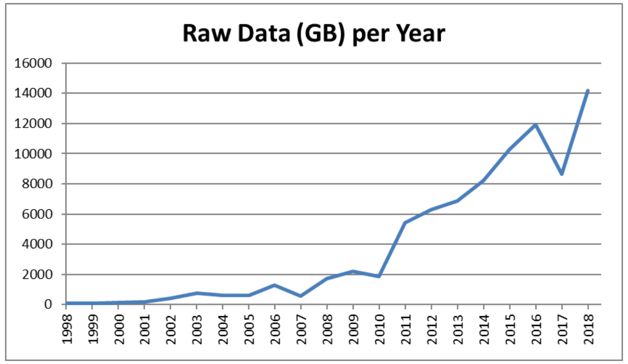


Data

- Volumes
- Fast/slow access
- Metadata
- Preservation, curation
- Policy







UKRI Science and Technology Facilities Council

ISIS

Diverse demands for computing: Examples

Macromolecular Crystallography

- High throughput beamlines requiring near real-time feedback on data collected
- Typical collecting 300 to 900 data sets per day,
- Data range: 5TB to 35TB per day RAW data, 2.5TB to 17TB per day processed data

Imaging Beamline

- 10 to 200 datasets per day
- Data range: 0.5TB to 20TB per day (technique dependant and detector dependant)

Electron Microscopy (eBIC – Life Sciences)

- 1 dataset collected per 48 hours
- Ranges: 4TB to 17TB per day

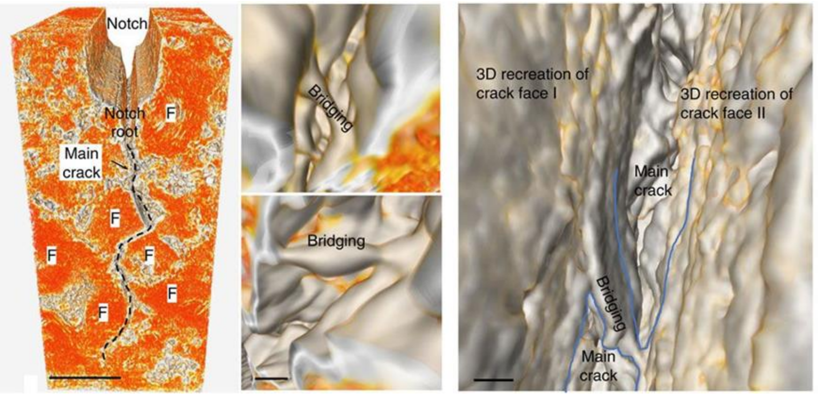
Electron Microscopy (ePSIC – Physical Sciences)

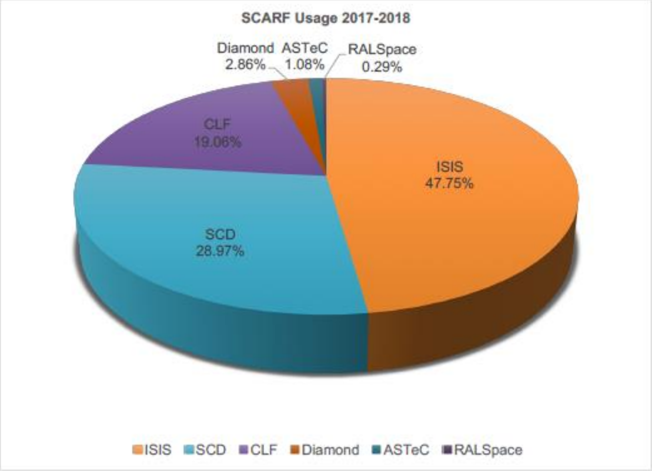
- 20 to 300 datasets per day
- Data range few GB’s to 100’s GB per day

diamond

Computing infrastructure

- Hardware – at facility or elsewhere (e.g. SCD Scarf)
- Fast/slow access
- Cloud (e.g. iDAaaS)
- Support for modelling/theory?
- Networking (including WiFi, Eduroam)





UKRI Science and Technology Facilities Council ISIS

Diverse demands for computing: Infrastructure

Increasing data rates and data volume require ever increasing computational infrastructure to be deployed to support the demand

- Diamond currently operates and consumes:
- Two onsite data centres
- ~6000 X86 cores
- ~200GPUs
- ~16PB of high performance storage (GPFS) for data acquisition
- Utilises offsite IRIS infrastructure for a few pioneering projects and exploring other ‘cloud’ services particularly for post-visit analysis
- Retains experimental data on-disk for 30 days
  - High data rate experiments are impacted most by this retention period in managing their workflow for data analysis. After 30 days data has to be restored from archive for further processing.
  - Some low data rate beamlines in the ‘long tail’ can be afforded the luxury of keeping data online ‘indefinitely’.
  - Overall data management has not kept up with the growth in data volumes. Poor data management is an increasing risk to data and timely analysis of data than just lacking computational resource to process data.

Diamond Future Cloud

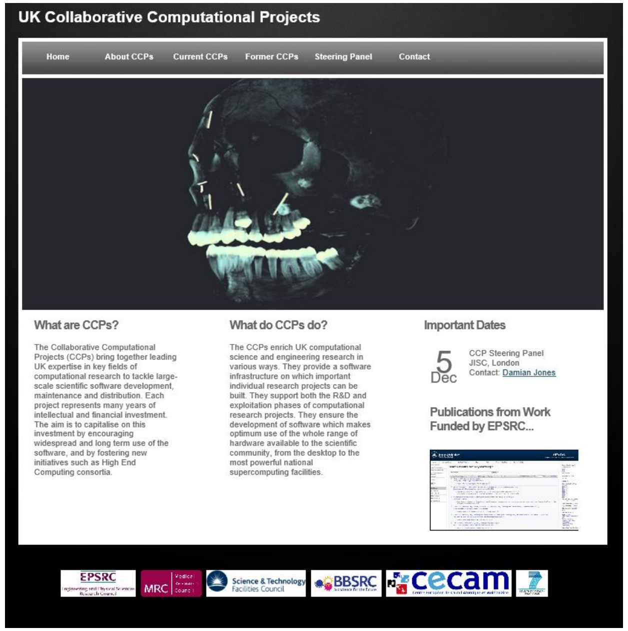
- **Offline / Post Processing** – enabling users to process data sets beyond the std. 40 day period ‘**easily’**
  - Need to transfer data from DLS to service (driven by DLS, DLS requirement to move offsite certain types of data processing)
  - Make ‘Diamond Software stack’ easily available e.g. via Docker/singularity

- **Burst computing** – ‘top-up’ backend infrastructure for Diamond
  - Transparent to end users
  - Only works for defined, relatively small (depends on target site) datasets.
  - Time – dependant analysis

- **Hosting** analysis services based around **Jupyter Hub** notebooks – user driven/user customisable
  - Need to transfer data from DLS to service (on-demand user driven service?)
  - Scalable infrastructure need

Software development

- Full experiment lifecycle: Preparation, Experiment, Reduction, Analysis, Visualisation, Modelling
- Transition from ‘lone amateur’ to ‘professional team’’ e.g. Mantid
- Collaborative Computational Projects



UK Collaborative Computational Projects

Home | About CCPs | Current CCPs | Former CCPs | Steering Panel | Contact

What are CCPs?

The Collaborative Computational Projects (CCPs) bring together leading UK expertise in key fields of computational research to tackle large-scale scientific software development, maintenance and distribution. Each project represents many years of intellectual and financial investment. The aim is to capitalise on this investment by encouraging widespread and long term use of the software, and by fostering new initiatives such as High End Computing consortia.

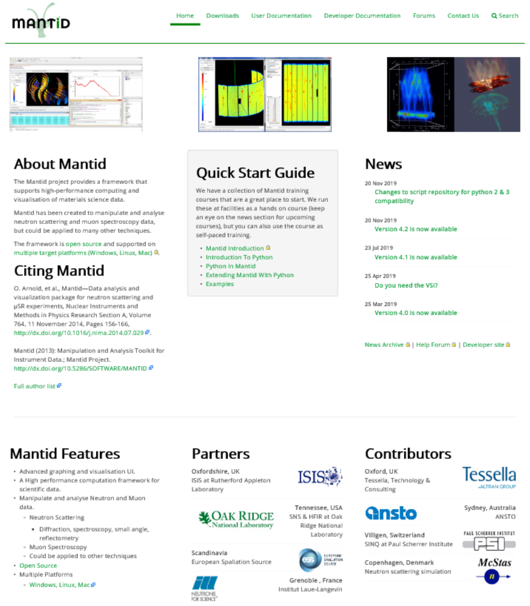
What do CCPs do?

The CCPs enrich UK computational science and engineering research in various ways. They provide a software infrastructure on which important individual research projects can be built. They support both the R&D and exploitation phases of computational research projects. They ensure the development of software which makes optimum use of the whole range of hardware available to the scientific community, from the desktop to the most powerful national supercomputing facilities.

Important Dates

5 Dec
CCP Steering Panel
JISC, London
Contact: Damian Jones

Publications from Work Funded by EPSRC...



MANTID

Home | Downloads | User Documentation | Developer Documentation | Forums | Contact Us | Search

About Mantid

The Mantid project provides a framework that supports high-performance computing and visualisation of materials science data.

Mantid has been created to manipulate and analyse neutron scattering and muon spectroscopy data, but could be applied to many other techniques.

The framework is open source and supported on multiple target platforms (Windows, Linux, Mac).

Citing Mantid

O. Arnold, et al., Mantid—Data analysis and visualization package for neutron scattering and μSR experiments, Nuclear Instruments and Methods in Physics Research Section A, Volume 764, 11 November 2014, Pages 156-166, http://dx.doi.org/10.1016/j.nima.2014.07.029.

Mantid (2013): Manipulation and Analysis Toolkit for Instrument Data.; Mantid Project. http://dx.doi.org/10.5286/SOFTWARE/MANTID

Full author list

Quick Start Guide

We have a collection of Mantid training courses that are a great place to start. We run these at facilities as a hands on course (keep an eye on the news section for upcoming courses), but you can also use the course as self-paced training.

- Mantid Introduction
- Introduction To Python
- Python In Mantid
- Extending Mantid With Python
- Examples

News

20 Nov 2019
Changes to script repository for python 2 & 3 compatibility

20 Nov 2019
Version 4.2 is now available

23 Jul 2019
Version 4.1 is now available

25 Apr 2019
Do you need the VSI?

25 Mar 2019
Version 4.0 is now available

News Archive | Help Forum | Developer site

Mantid Features

- Advanced graphing and visualisation UI.
- A High performance computation framework for scientific data.
- Manipulate and analyse Neutron and Muon data.
  - Neutron Scattering
    - Diffraction, spectroscopy, small angle, reflectometry
  - Muon Spectroscopy
  - Could be applied to other techniques
- Open Source
- Multiple Platforms
  - Windows, Linux, Mac

Partners

Oxfordshire, UK
ISIS at Rutherford Appleton Laboratory

Tennessee, USA
SNS & HFIR at Oak Ridge National Laboratory

Scandinavia
European Spallation Source

Grenoble , France
Institut Laue-Langevin

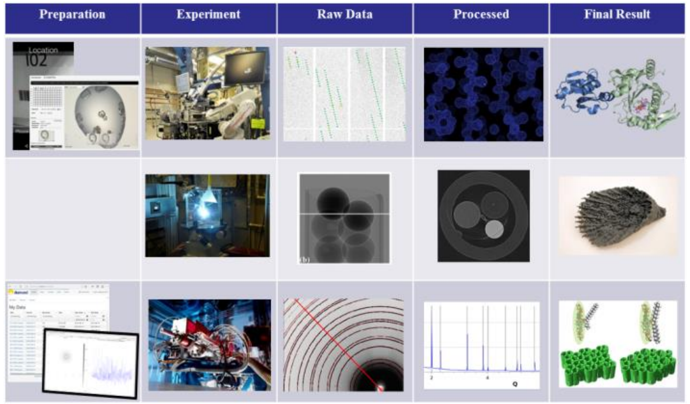
Contributors

Oxford, UK
Tessella, Technology & Consulting

Sydney, Australia
ANSTO

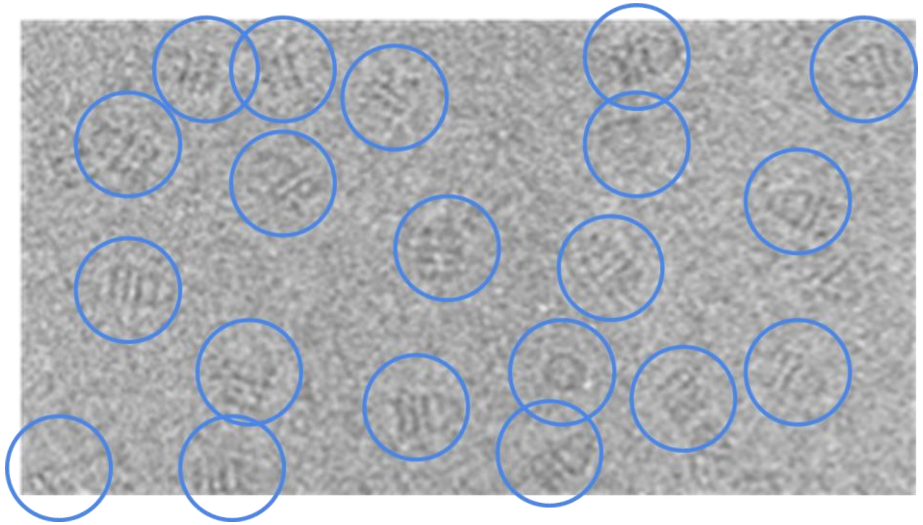
Villigen, Switzerland
SINQ at Paul Scherrer Institute

Copenhagen, Denmark
Neutron scattering simulation

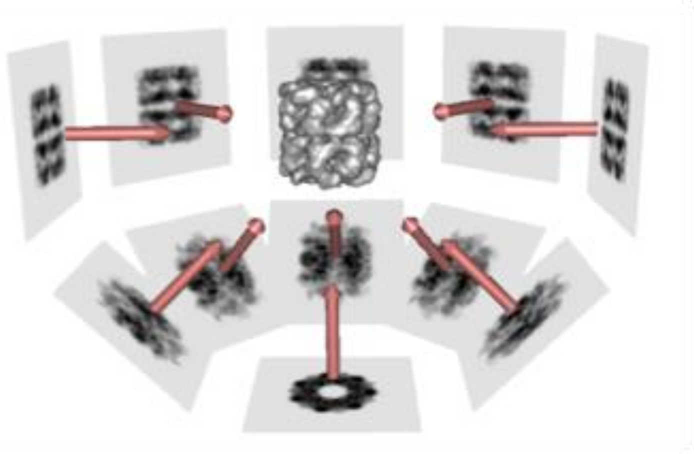


Machine Learning

- ‘Early days’ for facilities
- Imaging was an early adopter
- Exploring many other possibilities
- Data quality/relevance screening
- ‘Downstream’ databases?



CCP-EM



Science and Technology Facilities Council

ISIS

User expectations

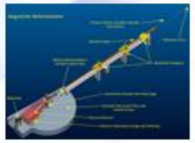
- More … and more … and more …

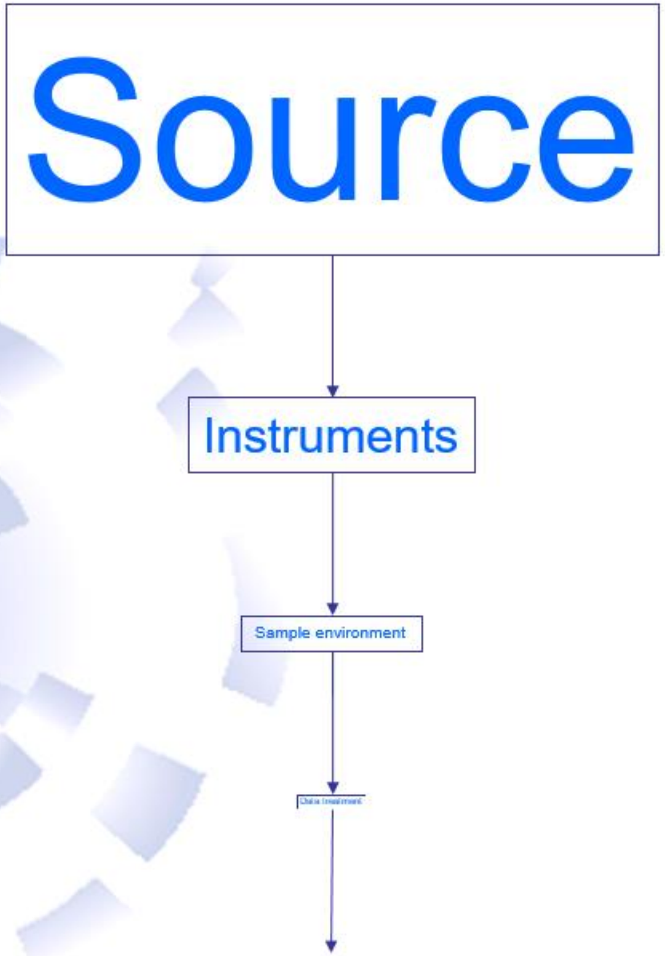
Science and Technology Facilities Council

ISIS

Skills

- Recruitment, training and retention …





Source
Instruments
Sample environment
Data treatment