Science & Technology Facilities Council

The ICAT Roadmap & STFC activities

Catherine Jones for Stuart Pullinger

Software Engineering Group Leader

20th March, 2018

ICAT tool set

The ICAT software is

- **ICAT Metadata Catalogue**: implementation of data model
- **ICAT Data Service**: enables data retrieval from storage
- **TopCat Web interface**: location tool

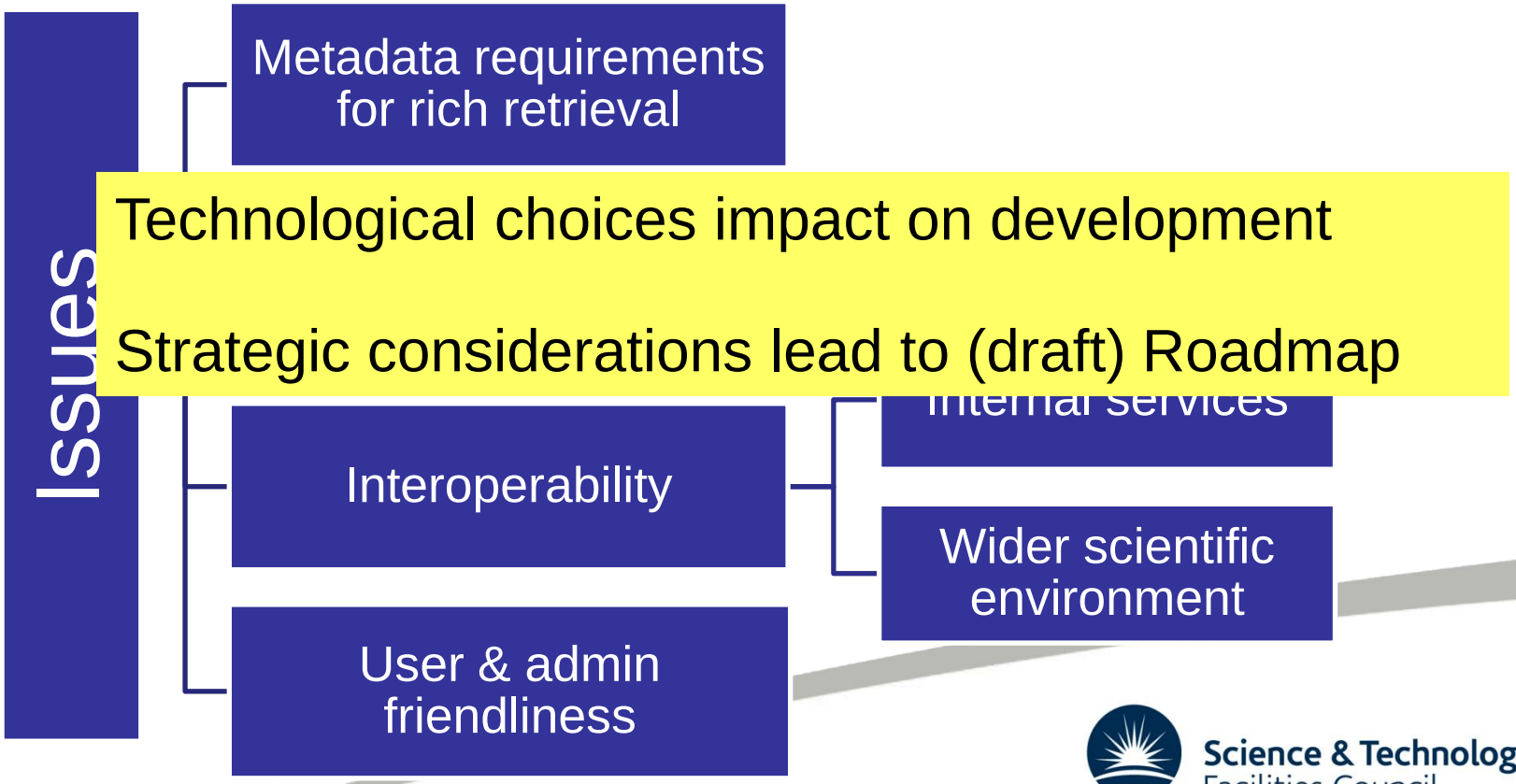
Current technology : JavaEE; RDMS & AngularJS

Open Source project: http://icatproject.org

Science & Technology Facilities Council

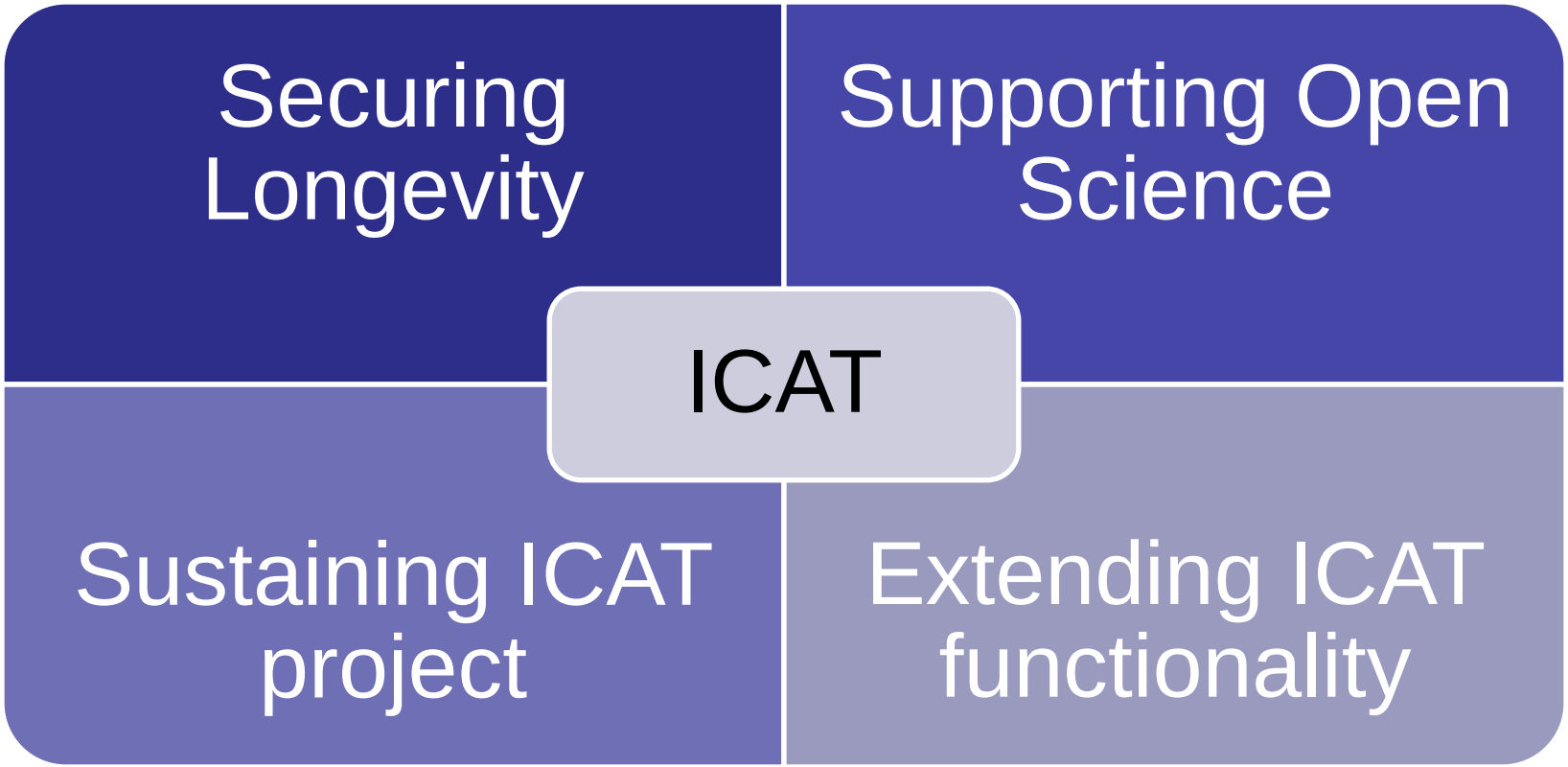
Status check

- Active ICAT community
- Requirements and technology change



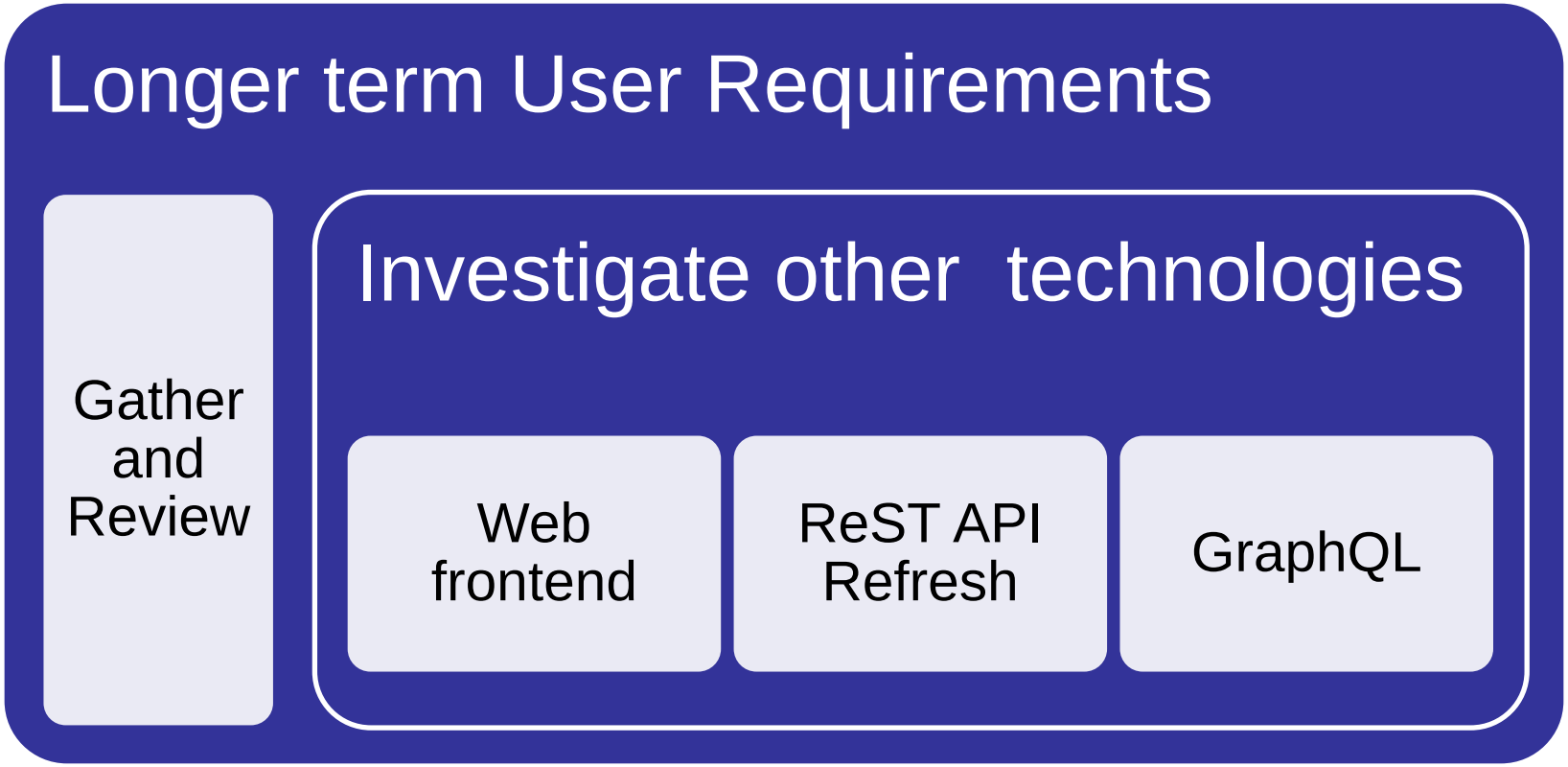
Science & Technology Facilities Council

Roadmap: Four themes



Science & Technology Facilities Council

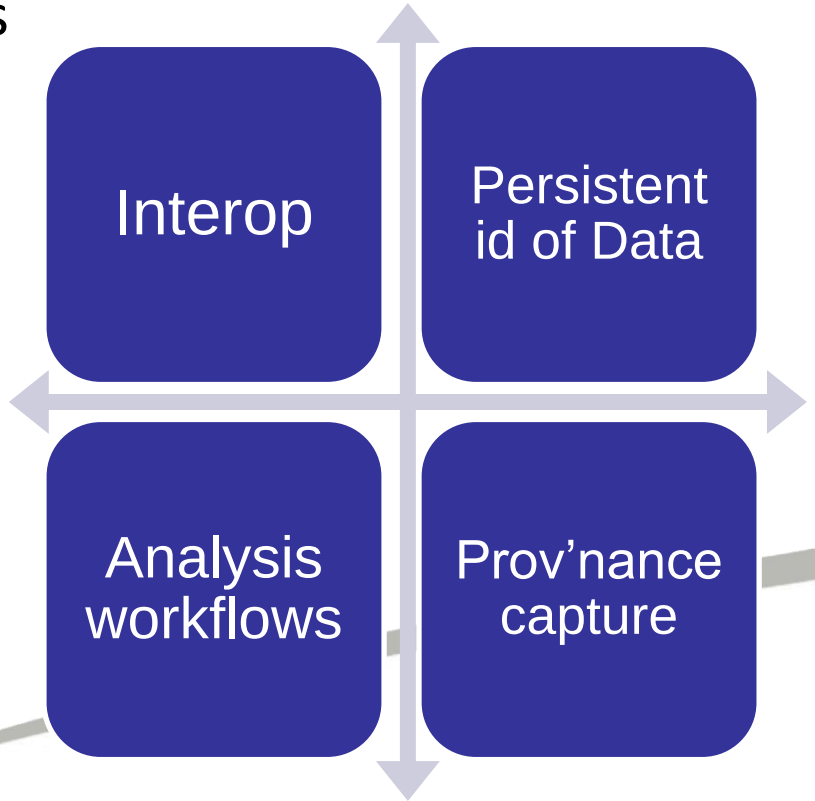
Securing Longevity



Science & Technology Facilities Council

Supporting Open Science

- Activities that enable data held in ICATs to comply with FAIR principles
- Encourage discovery & effective reuse
- Remembering that experimental data not the only data for our users



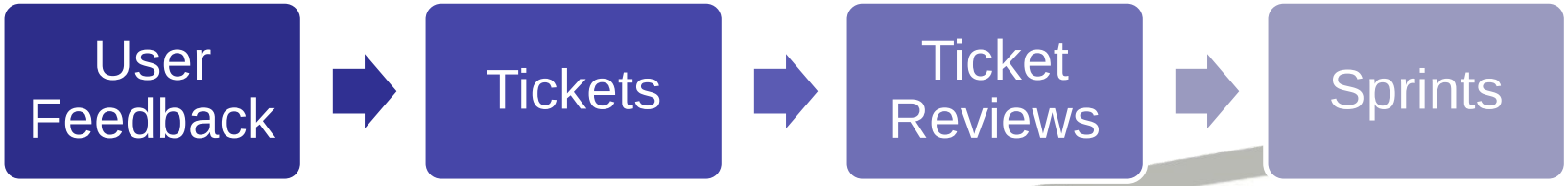
Science & Technology Facilities Council

Sustaining the ICAT Project

- Single to Multi-developer
  - Policies
  - Extending and enhancing automated testing
  - Metrics for code coverage
  - Docs & Web

Extending ICAT Functionality

- Key themes:
- **SCHEMA changes**
  - Support data publication
  - Considerations for performance improvements
- **Richer Metadata with existing constraints**
  - Going beyond key-value pairs
  - Allowing Arrays, hierarchies & JSON datatypes



Science & Technology Facilities Council

Next steps

- Confirmation and prioritisation by Steering Committee
- Strategic focus on user requirements and landscape survey

Science & Technology Facilities Council

STFC activities on DAaaS

Credit: Frazer Barnsley, Alastair Duncan & Brian Ritchie

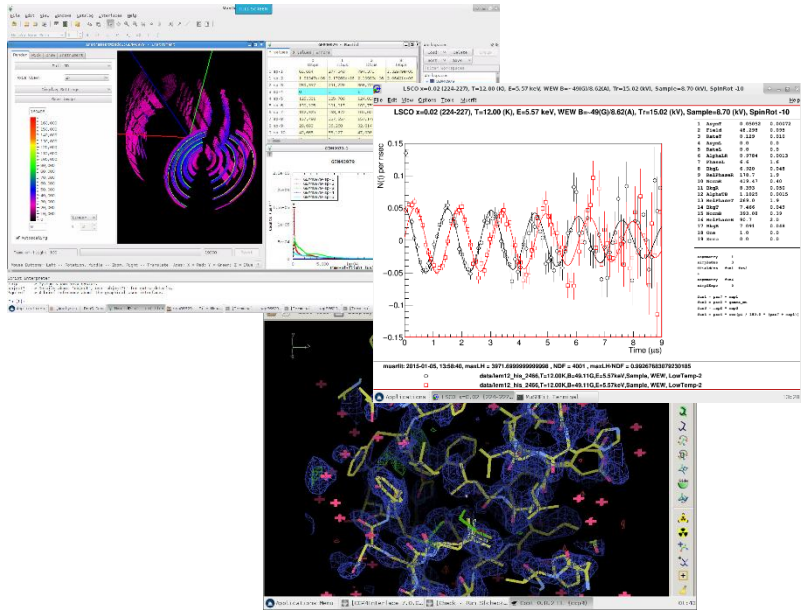
Science & Technology
Facilities Council

DAaaS projects & related activities

- ICAT Job Portal
  - Enabling Octopus Facility (Lasers) to manage data and run HPC analysis jobs in a batch modes within ICAT functionality
  - Using Singularity containers on our local cluster for analysis
- Data/Publication linking
  - Capturing relationships between research objects.
  - Jupyter: investigating use as another entry point to DAaaS
- IDAaaS: Software, Compute & Data
  - Enabling ISIS users to analyse their data locally to the data, in a friendly fashion for non-computing experts:
  - Automated deployment analysis s/w from Jenkins

Analysis Environments

- Remote desktop to VMs
- Customised to specific types of analysis
- Ready to go instantly
- Pre / in / post experiment analysis



Science & Technology Facilities Council

Science & Technology Facilities Council

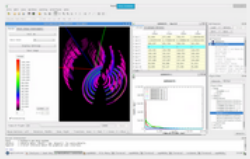
Home About Contact Help Admin

4 items Logout (Mr Frazer Barnsley)

Create Analysis Environments

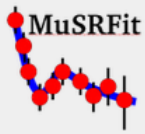
Excitations

This environment has been tailored for the analysis of in-elastic neutron scattering data from the ISIS Excitations Group. It comes complete with Mantid, Matlab, Horace and MSlice as well as access to the ISIS Experiment Archive.



Excitations Large

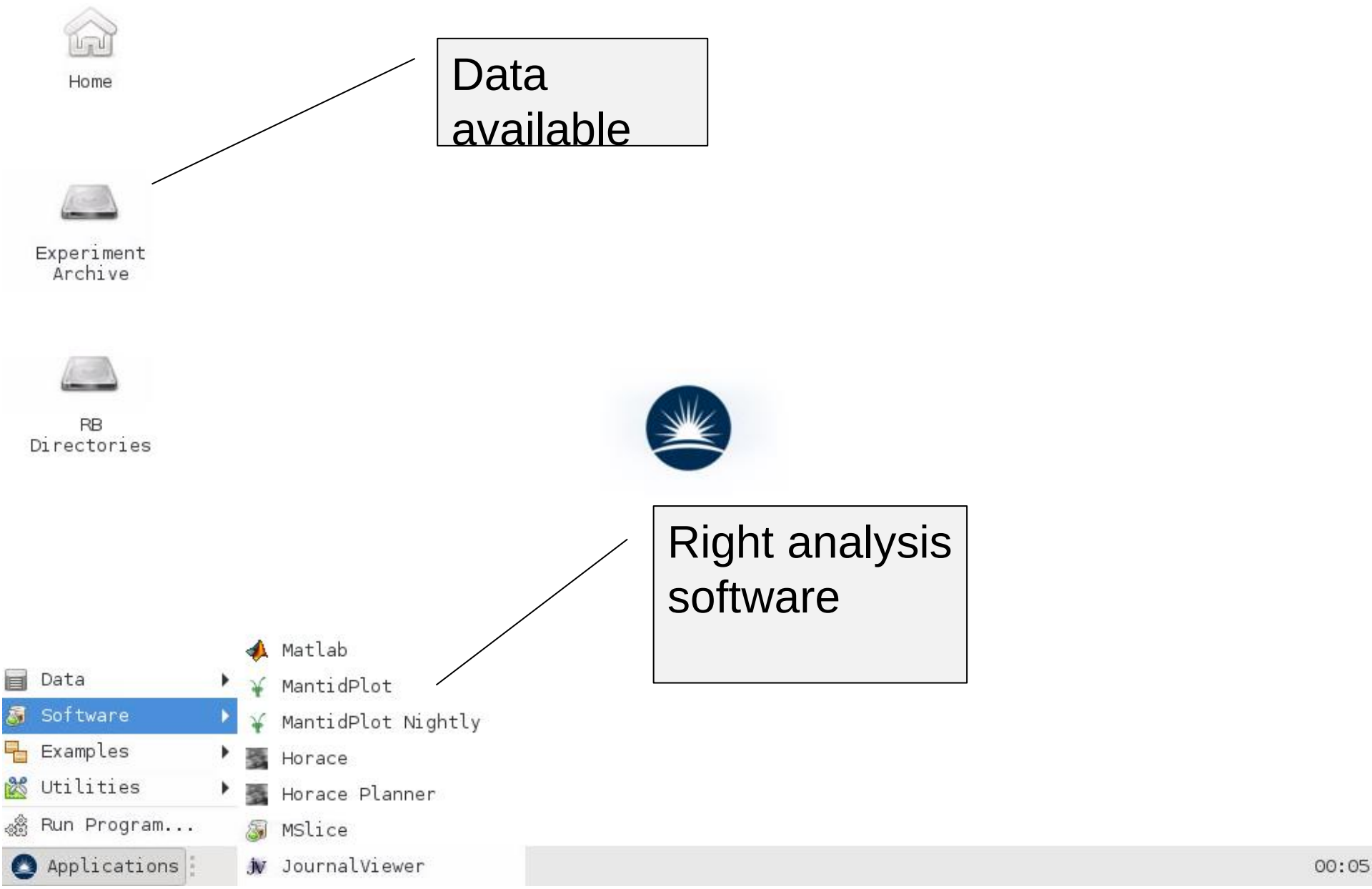
This environment has been tailored for the analysis of in-elastic neutron scattering data from the ISIS Excitations Group. It comes complete with Mantid, Matlab, Horace and MSlice as well as access to the ISIS Experiment Archive. This environment has been supplied with extra resources: 32GB RAM, 8 cores.



Muon

This environment has be setup to test the MuSRFit software.

Create Close


Home
Data available
Experiment Archive
RB Directories
Right analysis software
Matlab
Data
MantidPlot
Software
MantidPlot Nightly
Examples
Horace
Utilities
Horace Planner
Run Program...
MSlice
Applications
JournalViewer
00:05

Virtual Machine Manager Motivations

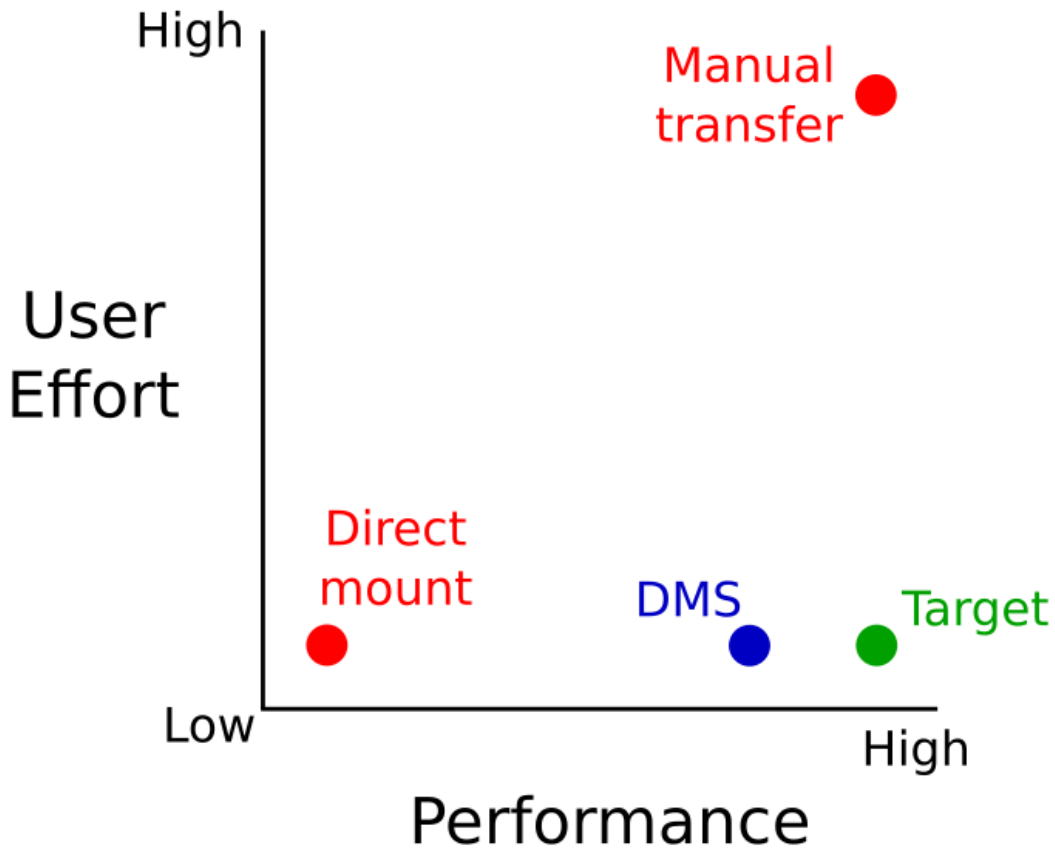
- Currently analysis environments only accessible through the TopCAT frontend
  - Open it up to other frontends
- Make use of resources at universities, other institutions, commercial clouds
- Bring ‘compute’ to the user / data

Requirements

- Provide API that allows clients to acquire different type analysis environments
- Cloud agnostic
  - Openstack, Azure ...
- Allow selection based on:
  - Location
  - Resources (CPU, RAM)
  - GPUs
- Fast – analysis environments available in seconds

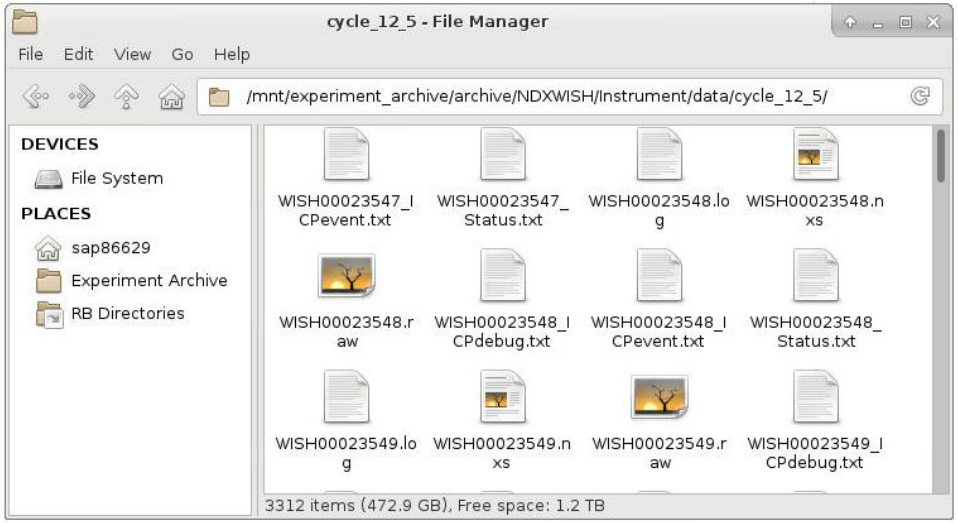
Data Management Service Motivations

- Simplify data access for users



DMS Requirements

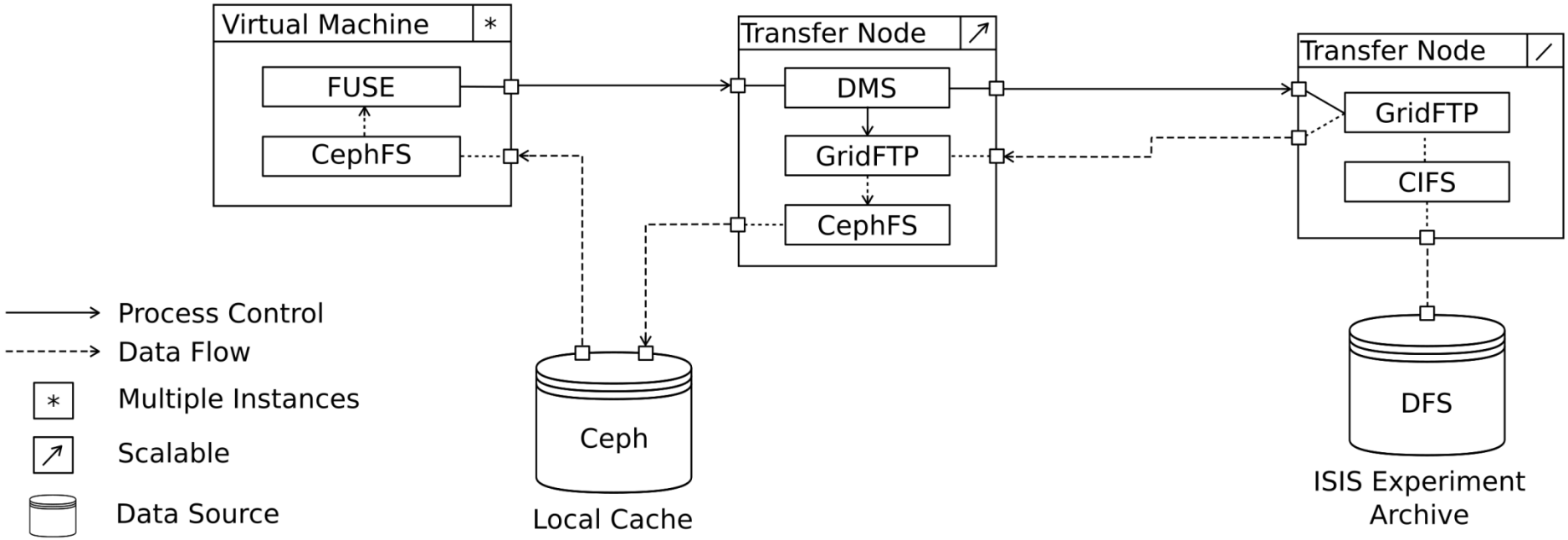
- Filesystem view of data for the user
- Passive - independence of storage resources
- Scalable for large data volumes
- Transfers over long distances



Design

- 2 main parts
  - FUSE client
  - DMS service

- Plus GridFTP machines



Achievements

- Virtual Machine Manager
  - API, cloud abstraction layer, pool manager complete
  - On-going development on allocation logic
  - Basic functional tests run against Openstack

- Data Movement Service
  - Can handle ISIS experiment archive (read-only, shared data)
  - Private user data almost complete (home dirs)
  - To-do, shared private data (ACL rules)
  - Performance tuning

- On-going development to integrate with DAaaS

Contact details

- Stuart.pullinger@stfc.ac.uk : ICAT Project Manager
- Frazer.Barnsley@stfc.ac.uk : DAaaS Technical Lead

- Catherine.jones@stfc.ac.uk : Software Engineering Group Leader

Science & Technology Facilities Council