# The SRB service at STFC

## and the road to iRODS(?)

Roger Downing
Kevin O'Neill

iRODS Workshop, Lyon
1 Feb, 2009

# Science and Technology Facilities Council – STFC

Formed by combining CCLRC (labs) & PPARC (PP + astronomy funding)

We're ex-CCLRC, so you get our labs

# The mission of the STFC e-Science centre is:

to spearhead the exploitation of e-Science technologies throughout STFC's programmes, the research communities they support, and the national science and engineering base.

Currently, this is mostly through facilities and programmes with physical presences at the labs

29/01/09

**Science & Technology**
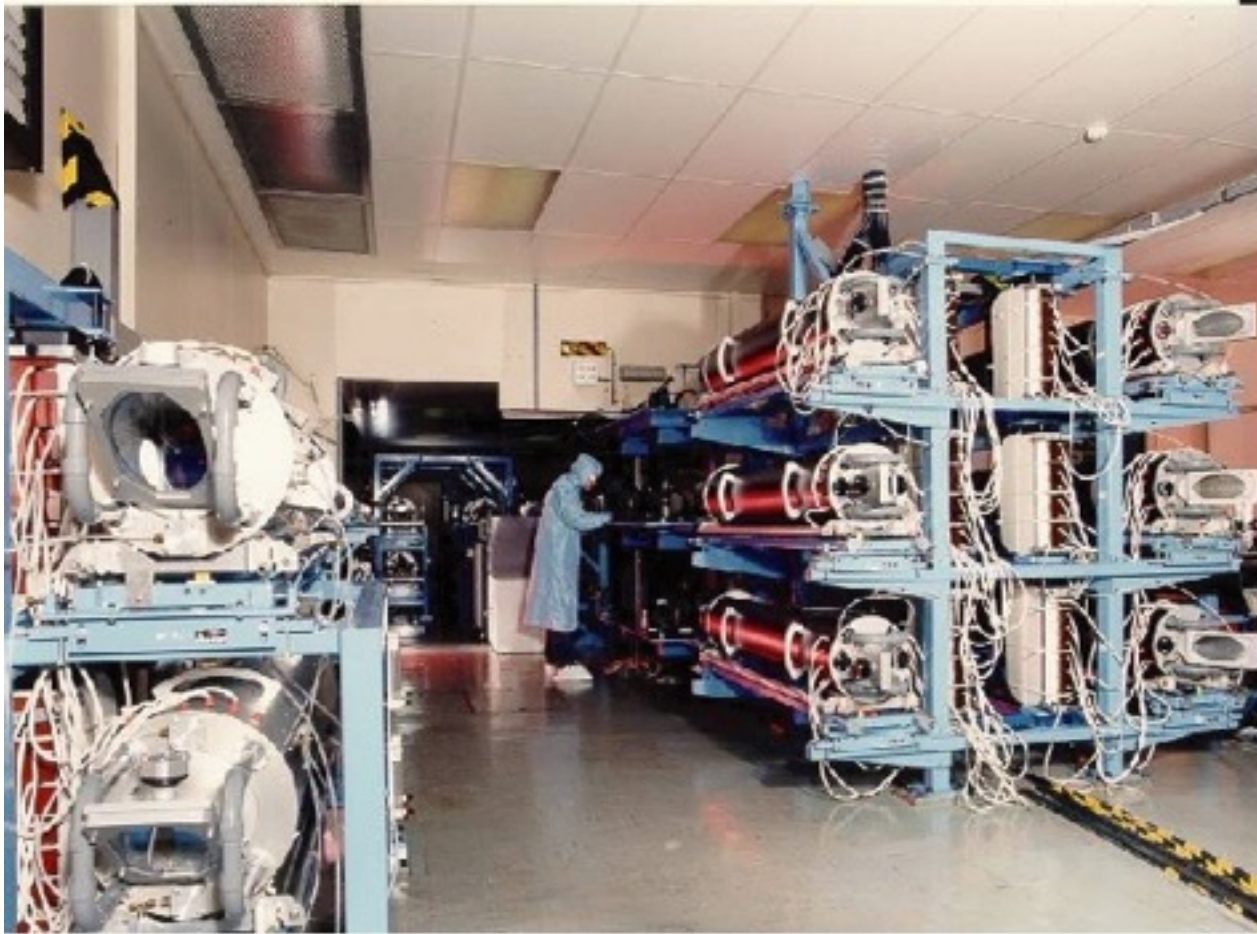Facilities Council

# STFC Facilities ISIS
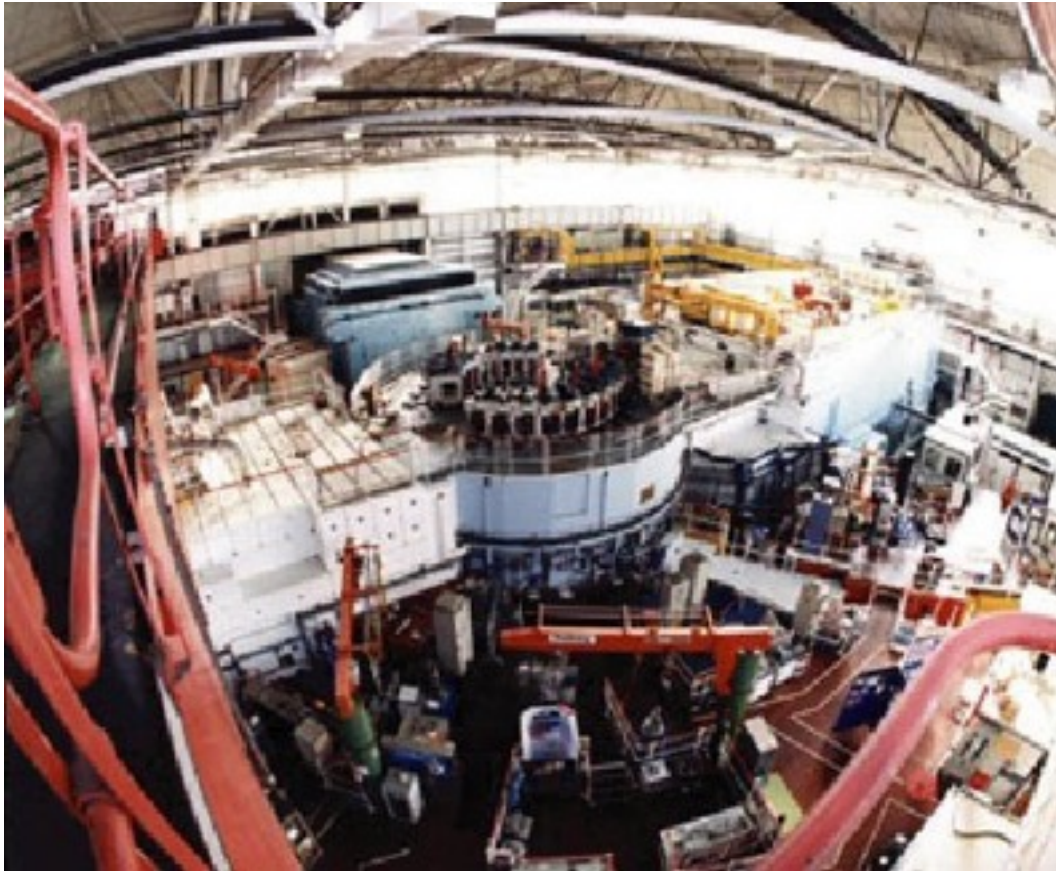
ISIS Neutron and Muon Facility

**Science & Technology**
Facilities Council

# STFC Facilities Central Laser Facility

Vulcan Petawatt Laser

Science & Technology
Facilities Council

# STFC Facilities – DLS



# Diamond Light Source

# The eScience Centre – SCARF

## *Scientific Computing Application Resource for Facilities*

To provide large scale computing with rapid access and turn round exclusively for users of CCLRC, its facilities, and diamond

- **256 AMD Opteron CPUs, 616GB RAM**
- **Parallel application focused**
- **16TB Filespace**
- **Free to STFC and STFC's users**
- **Grid based access**
- **http://www.scarf.rl.ac.uk to apply**

**A Happy User (Dr Matthias Gutmann from ISIS) looking at the results from SCARF**

**Dr Peter Oliver working on the installation of SCARF**

Support transparent access through NGS interfaces

logy

# ...and computing initiatives like...

- National Grid Service

- e-HTPX

- LHC computing and Tier-1 data management

- Digital Curation Centre

- ...etc...

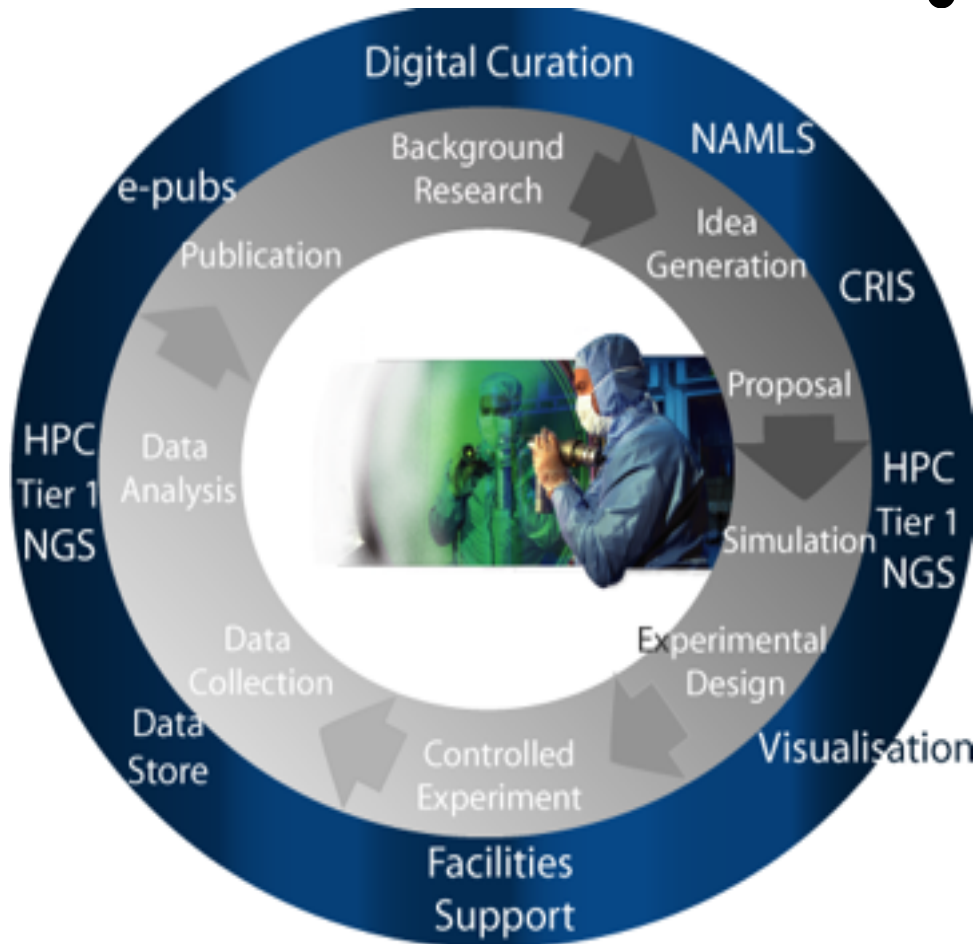**Science & Technology**
Facilities Council

# All this produces a lot of data...

- ...and it's no longer seen as "throw away"!
  - Even by the scientists producing it and the people funding it ;-)

- This all implies a change in our culture
  - Just as all the resources disappear

# We have a Cunning Plan…



- STFC e-Science infrastructure for the curation lifecycle, including (but not limited to):
  - Data storage
  - Data access
  - Data discovery
  - Metadata capture and management
  - Links to publications

29/01/09

**Science & Technology Facilities Council**

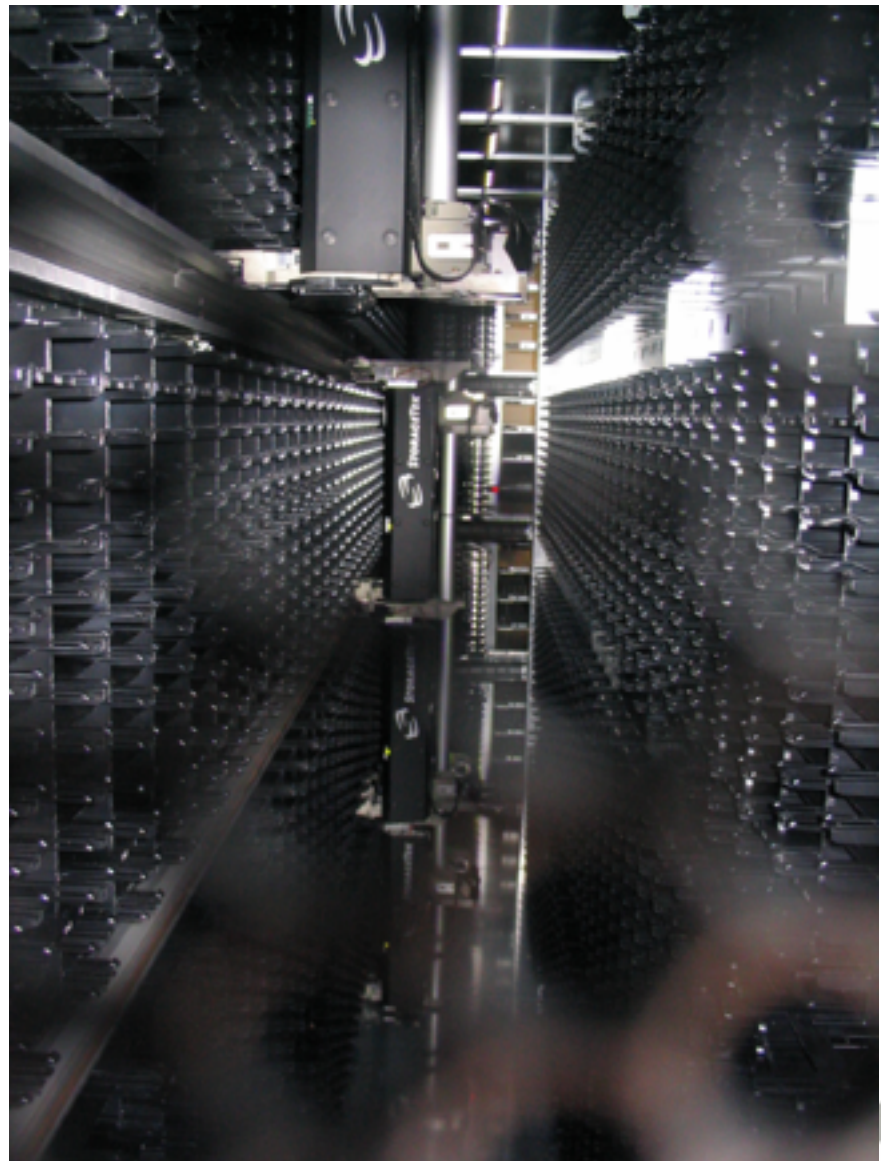**Science & Technology Facilities Council**
**e-Science**
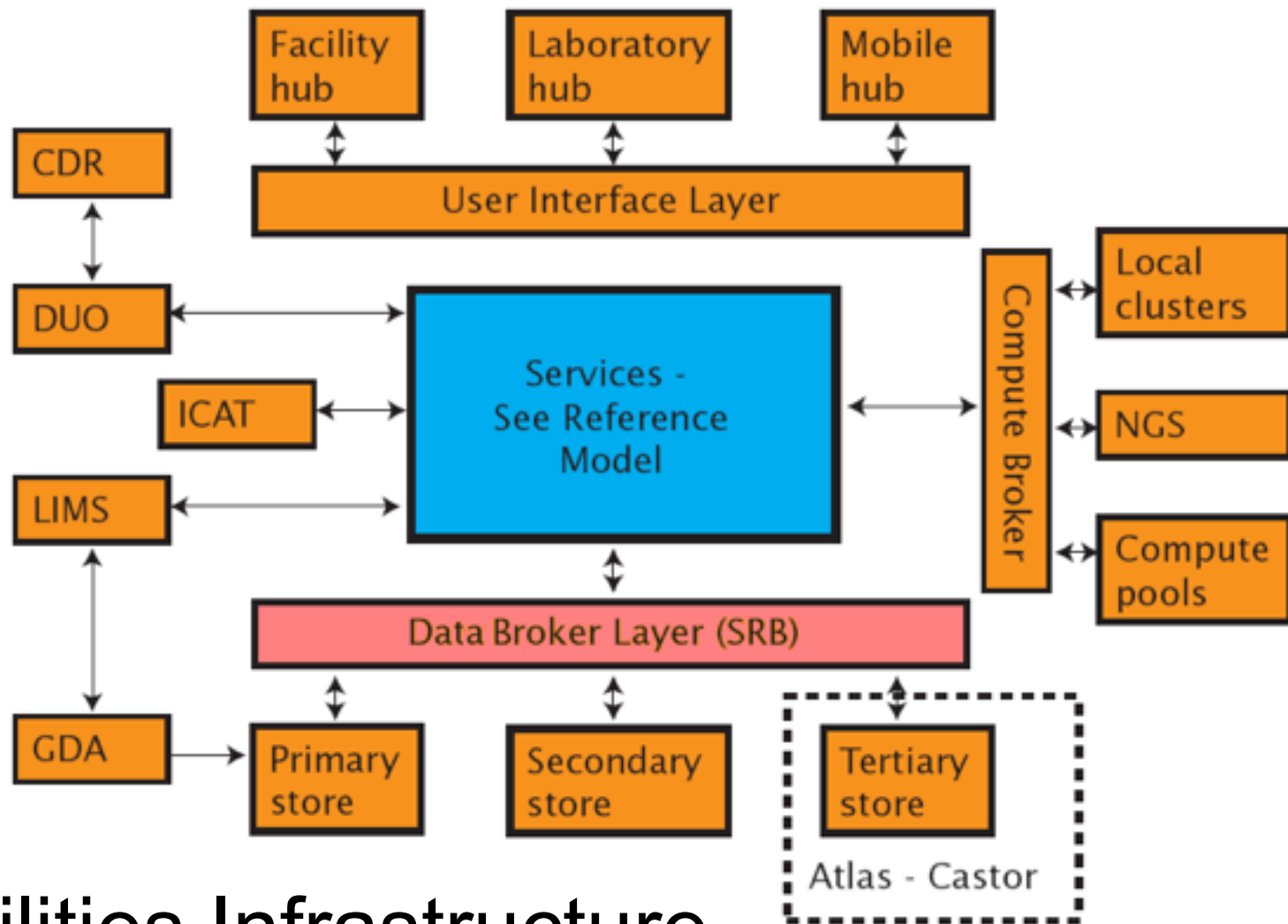
# 5 Petabytes of on line storage

# Atlas Petabyte DataStore

# 5 Petabytes of on line storage



Science & Technology
Facilities Council

# Facilities Infrastructure Architecture

# Main SRB-based services

- STFC facilities
  - Synchrotron Radiation Source (SRS)
  - Central Laser Facility (CLF)
  - ISIS Muon & Neutron Source
  - Diamond Light Source (DLS)
- External customers
  - Arts & Humanities Data Service (AHDS)
  - Biotechnology and Biological Sciences Research Council (BBSRC)
- BBSRC and DLS the most challenging
  - So we'll talk about them...

Science & Technology
Facilities Council

# BBSRC
## SRB as a "commercial" service

- BBSRC is the UK's lead funding agency for academic research and training in the non-clinical life sciences

- Data was held at individual institutes, and not available elsewhere

- Agreement  with BBSRC IT Service Centre to provide infrastructure to promote sharing

- Formal Service Level Agreement in place
  - Metrics to allow BBSRC to monitor compliance
  - Royalties to General Atomics

Science & Technology
Facilities Council

# BBSRC

- Service is successful
  - take up limited by bandwidth
  - is expected to the basis for advancing data curation practices in BBSRC



Science & Technology
Facilities Council
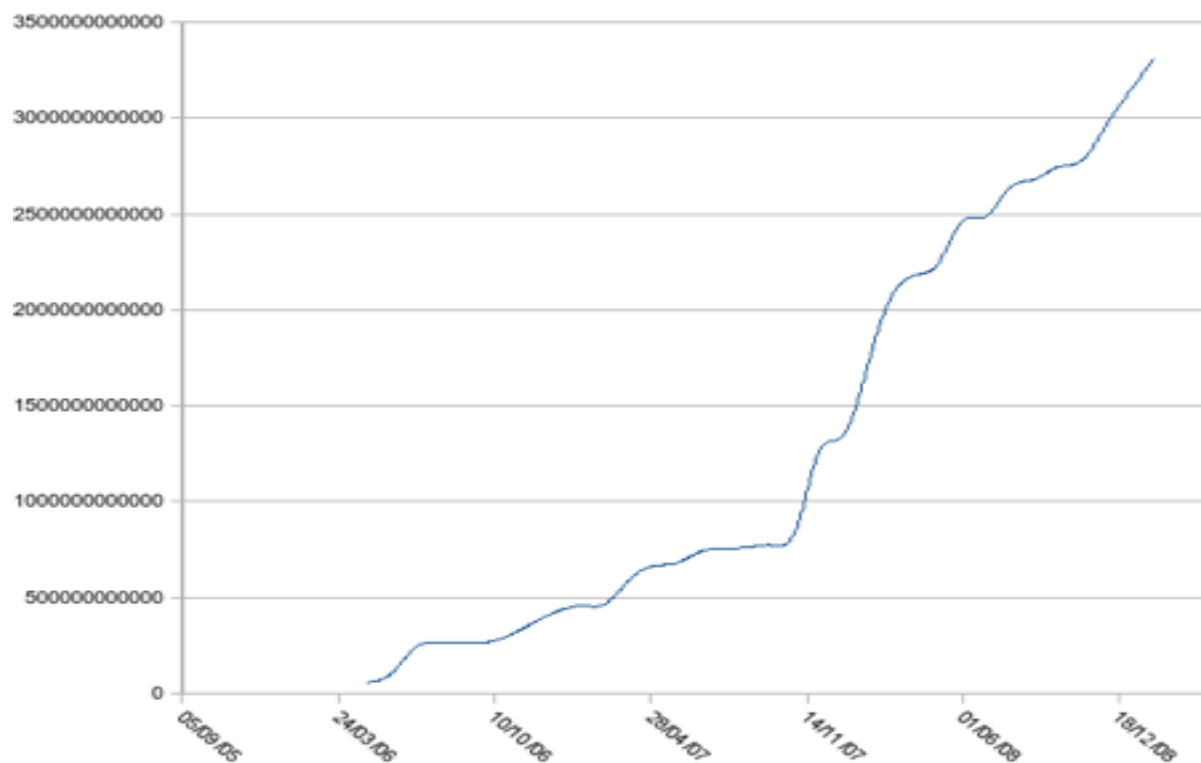
# BBSRC – General architecture

- Service is available to 14 BBSRC funded institutes with heterogeneous client platforms
- Each has local SRB server with disk resource uploading to a central BBSRC server
  - regularly run scripts uploading across the network to ADS
  - Extensive use of containers to make good use of limited bandwidth

Science & Technology
Facilities Council

# BBSRC system – key added features

- BBSRC designed metadata user interface

    - Most metadata inserted automatically, but some free-form fields to allow user additions

- Process control

    - Data logically in "packages" of a single upload session by a client

    - Resource tracker DB monitoring state of packages

Science & Technology
Facilities Council

# DLS

- Largest investment in UK science for at least 40 years
- Will soon be producing a Petabyte of data a year, and rising...
- Trying to get data managed as soon as possible!
- And all under a Service Level Agreement (SLA)

**Science & Technology**
Facilities Council

# DLS "Issues"

- Managing data from creation onwards
  - Data rate challenge
    - A lot produced in a short time
    - New detectors are producing even more
    - And DLS are deploying more detectors anyway
  - Large scale storage
    - Did we mention the data rates, and that they want to keep it for as long as possible?
  - Long-term archival
    - A process, not just a task

Science & Technology
Facilities Council

# DLS – Description of the process

# DLS –Challenges

- Staged storage
  - While we treat the SRB URIs as PIs, we still have to move the data between storage resources as it moves through the life cycle

- Workarounds for SRB limitations
  - Designation of a master copy
  - Assumption that all replicas are stored the same way
  - Lack of "connection pooling"

Science & Technology
Facilities Council

# More general problems encountered

- Performance
- DB issues
  - Examples
    - Many basic indices missing
    - Missing Primary/Foreign keys cripples many things…
    - No use of stored procedures/ functions

Science & Technology
Facilities Council

# More general problems encountered (2)

- Diagnostics
  - Logging
    - Log contents usually unhelpful
    - Log to syslog?
  - Debugging
    - Not always clear where the problem lies, errors often misleading

- Availability

Science & Technology
Facilities Council

# IRODS evaluation

- Many assuming that iRODS will be a natural successor to SRB

- But our plan is based around an infrastructure delivering function, not deploying technology in a project

- So we're
    - Treating SRB as our pilot
    - gathering our criteria, prior to testing
        - In so far as we can...

Science & Technology
Facilities Council

# IRODS evaluation criteria (1)

- This is a Work In Progress!
- Required functional features
  - Have interfaces for our storage resources (SRM interface?)
  - Container support
  - Migration path for end-user written code
    - So reproducing S-commands seamlessly would be good

29/01/09

Science & Technology
Facilities Council

# IRODS functional evaluation criteria (2)

- More required functional features
    - Replica management
    - Federation – ease and effectiveness
    - Able to cope with data rates
    - Scalability with many millions of files
    - Data input rate (RBUDP will be tried)

29/01/09

**Science & Technology**
Facilities Council

# IRODS evaluation criteria – more

- IRODS could be in place in a changing environment for decades. We need a product that is
  - Stable;
  - Robust;
  - Easy to maintain;
  - Free of licencing issues
  - Collaboratively developed to provide the effort

**Science & Technology**
Facilities Council

# IRODS evaluation criteria – more

- It also has to
  - Integrate as an equal into an existing production environment
    - Database services
    - Machine configurations (unixODBC?)
    - Security infrastructures
    - Supports established workflow mechanisms
    - Copes with multiple FTPs

**Science & Technology**
Facilities Council

# To sum up

- SRB serves us well
    - Learnt to avoid problem areas
    - But a lot of added code
- iRODS holds great promise

# But attention must be paid to long–term production usage issues

**Science & Technology**
Facilities Council

# Questions?

# Contacts

Roger Downing     – roger.downing@stfc.ac.uk
Kevin O'Neill    – kevin.o'neill@stfc.ac.uk

**Science & Technology**
Facilities Council