SRB – What and Why?

Kerstin Kleese van Dam

CCLRC - Daresbury and Rutherford Appleton Laboratories

k.kleese@dl.ac.uk

Introduction

More and more information is available today, it can be :

- **Random Information (e.g. news items)**
- **Scientific Data**
- **Commercial or Administrative Data**
- **Data about Data (metadata describing the content of the actual data)**

The information is generally available via/from:

Web-sites, Filesystems, Databases, Tape Libraries or on Paper and other none digital media.

Introduction (2)

How do you find the information:

Search Engines, Catalogue Systems or Hard Work (big bucket)

How do you evaluate the information:

Combine, Compare, Present

How do you manage the information:

Preservation, Sharing, Replicating, Transferring, Securing, Audit Trails

Where does SRB fit into this Scenario?

SRB - the Storage Resource Broker can:

- Integrate distributed, heterogeneous storage devices
- Make data access transparent for the user
- Helps to share, replicate, transfer and preserve data
- Supports international collaborations

SRB can not:

- Replace metadata catalogues
- Provide high level information services

How does SRB fit into a Grid Environment?

SRB can be used to:

Integrate data across various media

Integrate data across sites

Support distributed computing

SRB can be used:

For a particular site

In a research collaboration

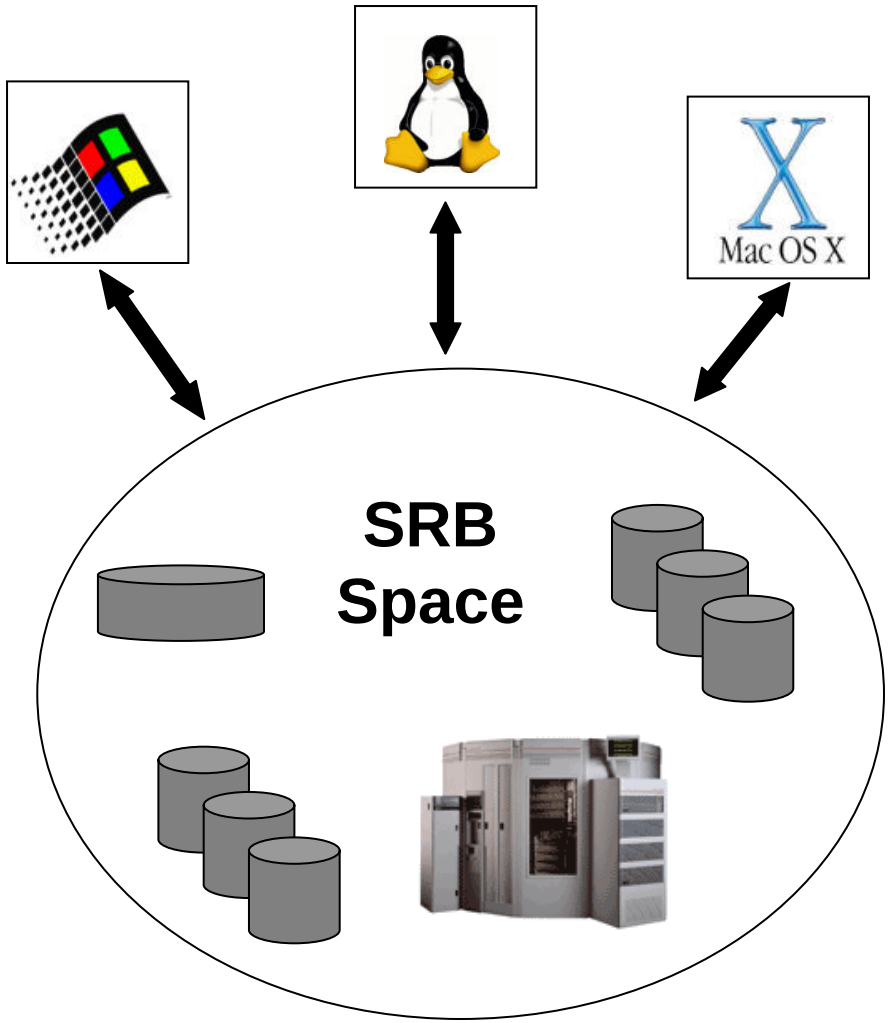
In a wider Grid community

Storage Resource Broker

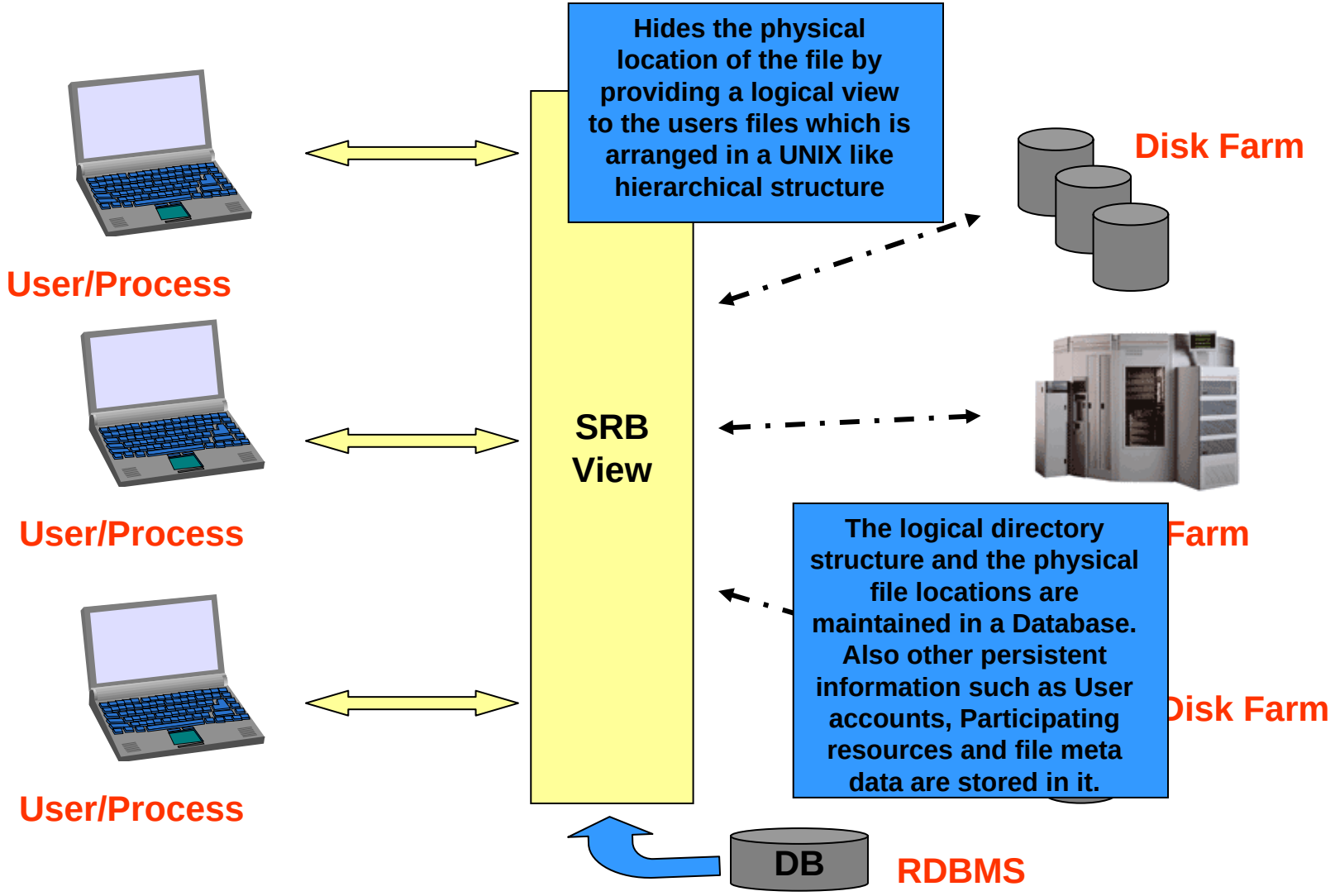
The SDSC Storage Resource Broker is a client server middleware that virtualizes data space by providing a unified view to multiple heterogeneous storage Resources over the network.

SRB, in conjunction provides a way to access data sets and resources based on their attributes rather than their names or physical locations.

The software sits in between users and resources and provides a storage service by managing users, file locations, storage resources and metadata information



SRB Logical Structure



General Facts about SRB

- **Developed by the San Diego Supercomputing Centre (SDSC) from the mid 1990's for the US governments' National Partnership for Advanced Computational Infrastructure (NPACI). http://www.npaci.edu/DICE/SRB/.**
- **Initial release 1997**
- **Latest version v3.2 - released July 2004**
- **In the US approximately 200TB of data are shared via SRB between 30 participating Universities.**
- **CCLRC runs currently 4 SRB services with approx. 50 TB of data, 1000000Files with 20 participating Universities and Research Institutes world wide.**

Interfaces to SRB

SRB User Interfaces

Users can access SRB via a number of well designed interfaces:

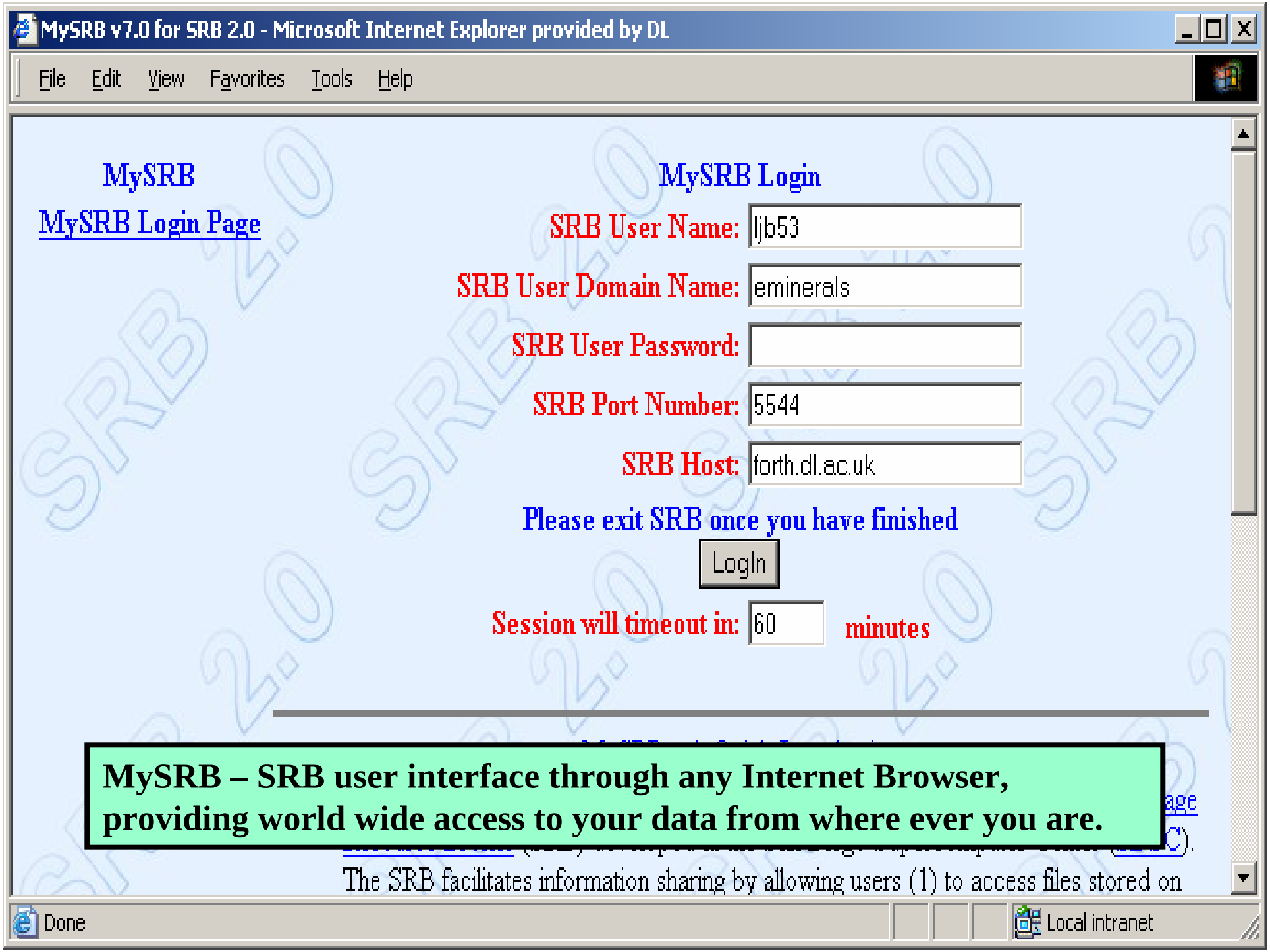
InQ – Windows based user interface supporting the full range of SRB commands

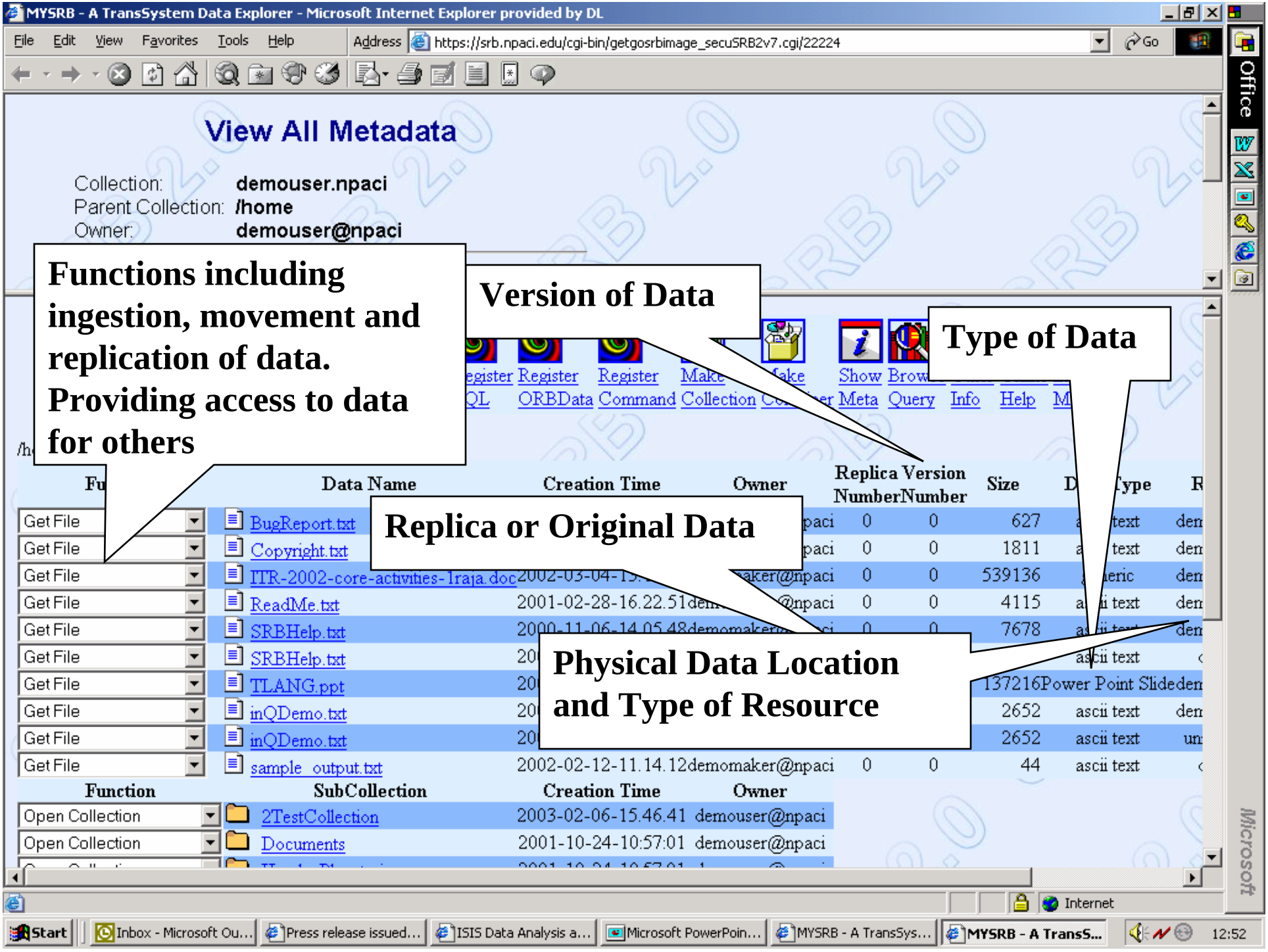
MySRB – Web based user interface allowing access to the users SRB data from anywhere in the world

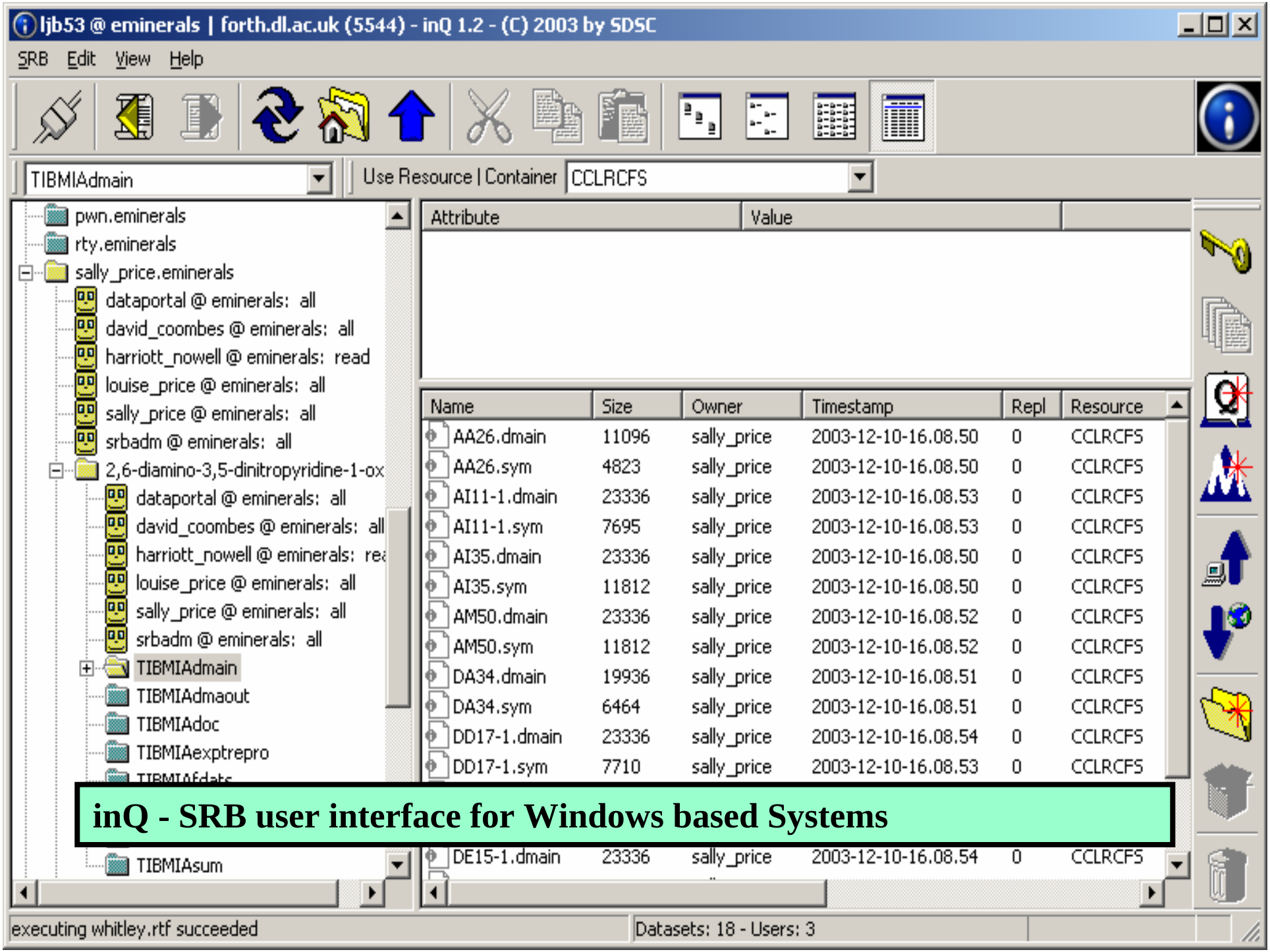
S-Commands – Command line interface to interact with SRB

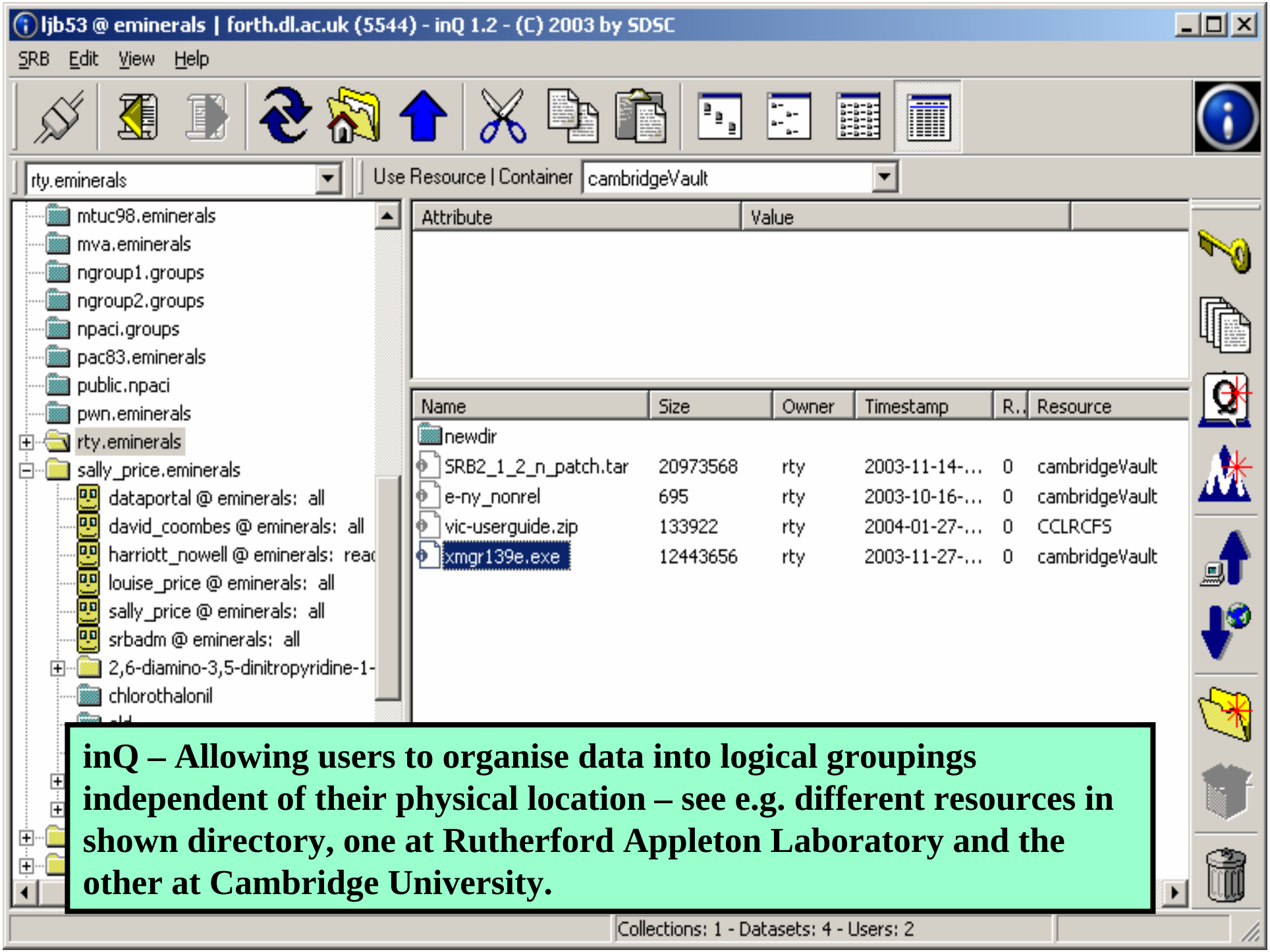
SRB-API for direct interaction with SRB from any programme, examples exist for e.g. Fortran, C, C++, VB, Perl and Python

There is also an Administration Interface

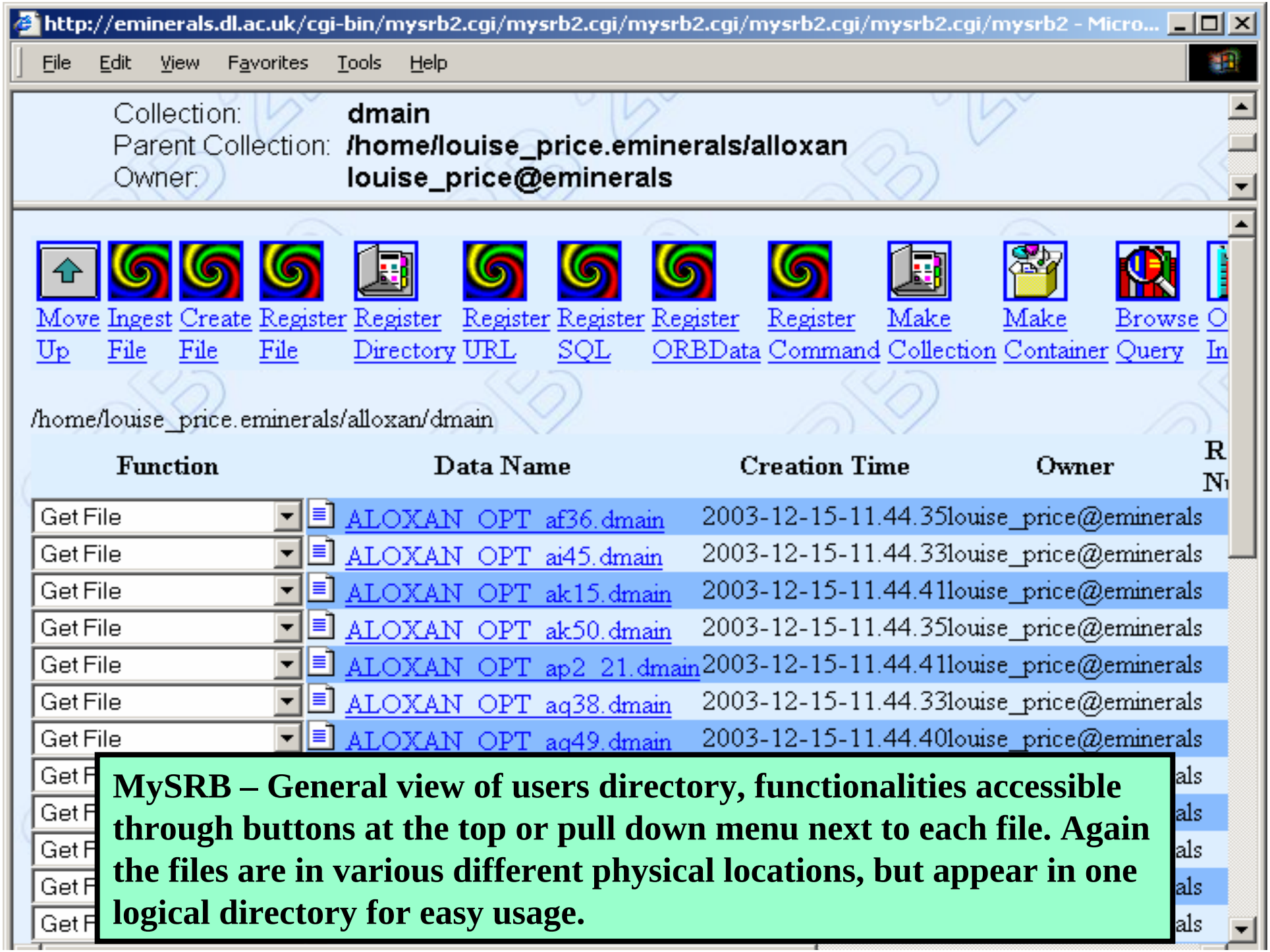

MySRB v7.0 for SRB 2.0 - Microsoft Internet Explorer provided by DL
File Edit View Favorites Tools Help
MySRB
MySRB Login Page
MySRB Login
SRB User Name: ljb53
SRB User Domain Name: eminerals
SRB User Password:
SRB Port Number: 5544
SRB Host: forth.dl.ac.uk
Please exit SRB once you have finished
LogIn
Session will timeout in: 60 minutes
MySRB – SRB user interface through any Internet Browser, providing world wide access to your data from where ever you are.
The SRB facilitates information sharing by allowing users (1) to access files stored on
Done
Local intranet


MYSRB - A TransSystem Data Explorer - Microsoft Internet Explorer provided by DL
https://srb.npaci.edu/cgi-bin/getgosrbimage_secuSRB2v7.cgi/22224
View All Metadata
Collection: demouser.npaci
Parent Collection: /home
Owner: demouser@npaci
Functions including ingestion, movement and replication of data. Providing access to data for others
Version of Data
Type of Data
Replica or Original Data
Physical Data Location and Type of Resource
Register ORBData
Register Command
Make Collection
Show Meta
Query
Info
Help
Data Name
Creation Time
Owner
Replica Number
Version Number
Size
Type
BugReport.txt
Copyright.txt
ITR-2002-core-activities-Iraja.doc
ReadMe.txt
SRBHelp.txt
SRBHelp.txt
TLANG.ppt
inQDemo.txt
inQDemo.txt
sample_output.txt
2002-02-12-11.14.12demomaker@npaci
Function
SubCollection
2TestCollection
2003-02-06-15.46.41 demouser@npaci
Documents
2001-10-24-10:57:01 demouser@npaci
Internet
Start
12:52

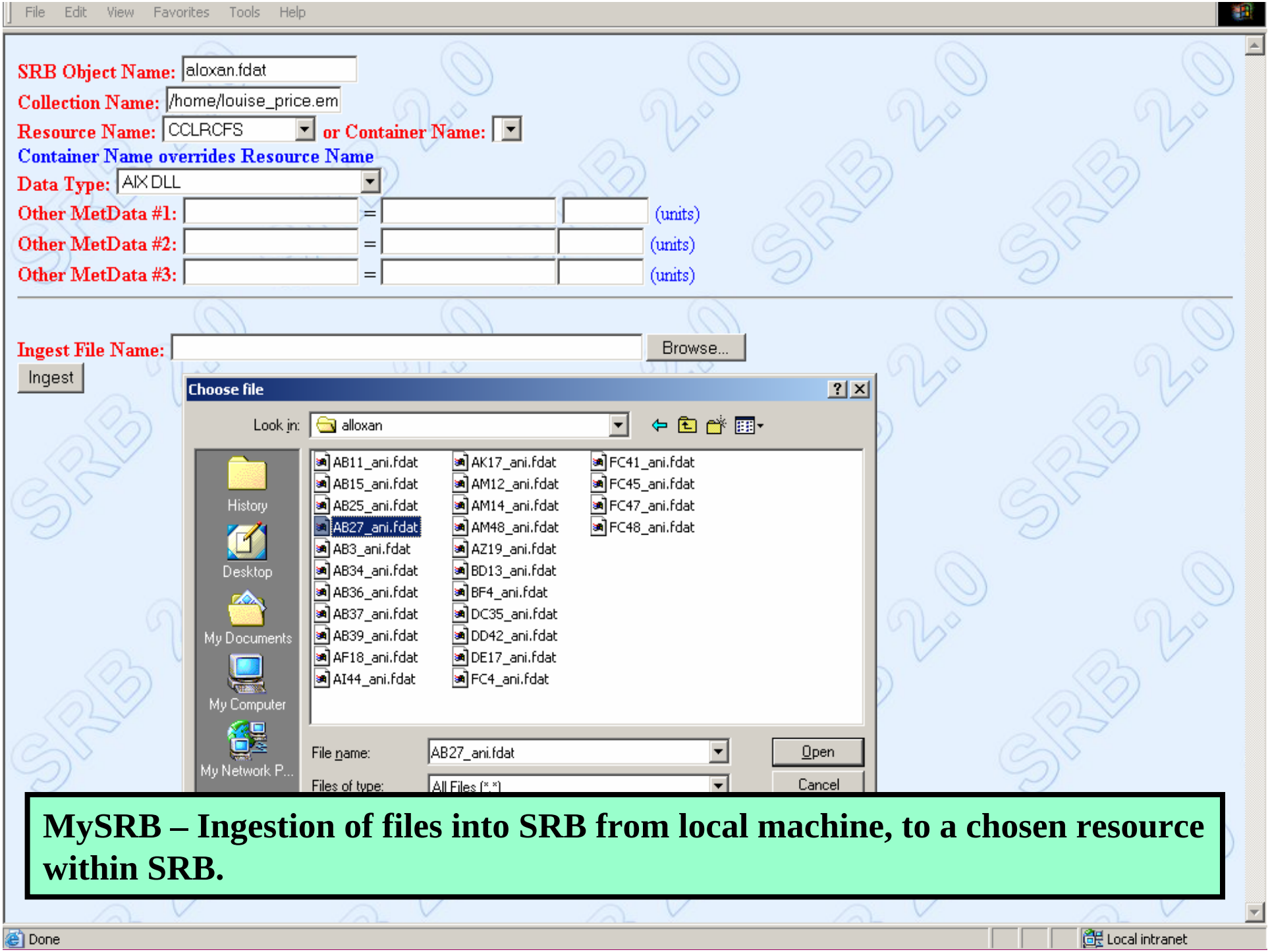

ljb53 @ eminerals | forth.dl.ac.uk (5544) - inQ 1.2 - (C) 2003 by SDSC
SRB Edit View Help
TIBMIAdmain
Use Resource | Container
CCLRCFS
pwn.eminerals
rty.eminerals
sally_price.eminerals
dataportal @ eminerals: all
david_coombes @ eminerals: all
harriott_nowell @ eminerals: read
louise_price @ eminerals: all
sally_price @ eminerals: all
srbadm @ eminerals: all
2,6-diamino-3,5-dinitropyridine-1-ox
dataportal @ eminerals: all
david_coombes @ eminerals: all
harriott_nowell @ eminerals: rea
louise_price @ eminerals: all
sally_price @ eminerals: all
srbadm @ eminerals: all
TIBMIAdmain
TIBMIAdmaout
TIBMIAdoc
TIBMIAexptrepro
TIBMIAsum
Attribute
Value
Name Size Owner Timestamp Repl Resource
AA26.dmain 11096 sally_price 2003-12-10-16.08.50 0 CCLRCFS
AA26.sym 4823 sally_price 2003-12-10-16.08.50 0 CCLRCFS
AI11-1.dmain 23336 sally_price 2003-12-10-16.08.53 0 CCLRCFS
AI11-1.sym 7695 sally_price 2003-12-10-16.08.53 0 CCLRCFS
AI35.dmain 23336 sally_price 2003-12-10-16.08.50 0 CCLRCFS
AI35.sym 11812 sally_price 2003-12-10-16.08.50 0 CCLRCFS
AM50.dmain 23336 sally_price 2003-12-10-16.08.52 0 CCLRCFS
AM50.sym 11812 sally_price 2003-12-10-16.08.52 0 CCLRCFS
DA34.dmain 19936 sally_price 2003-12-10-16.08.51 0 CCLRCFS
DA34.sym 6464 sally_price 2003-12-10-16.08.51 0 CCLRCFS
DD17-1.dmain 23336 sally_price 2003-12-10-16.08.54 0 CCLRCFS
DD17-1.sym 7710 sally_price 2003-12-10-16.08.53 0 CCLRCFS
DE15-1.dmain 23336 sally_price 2003-12-10-16.08.54 0 CCLRCFS
inQ - SRB user interface for Windows based Systems
executing whitley.rtf succeeded
Datasets: 18 - Users: 3



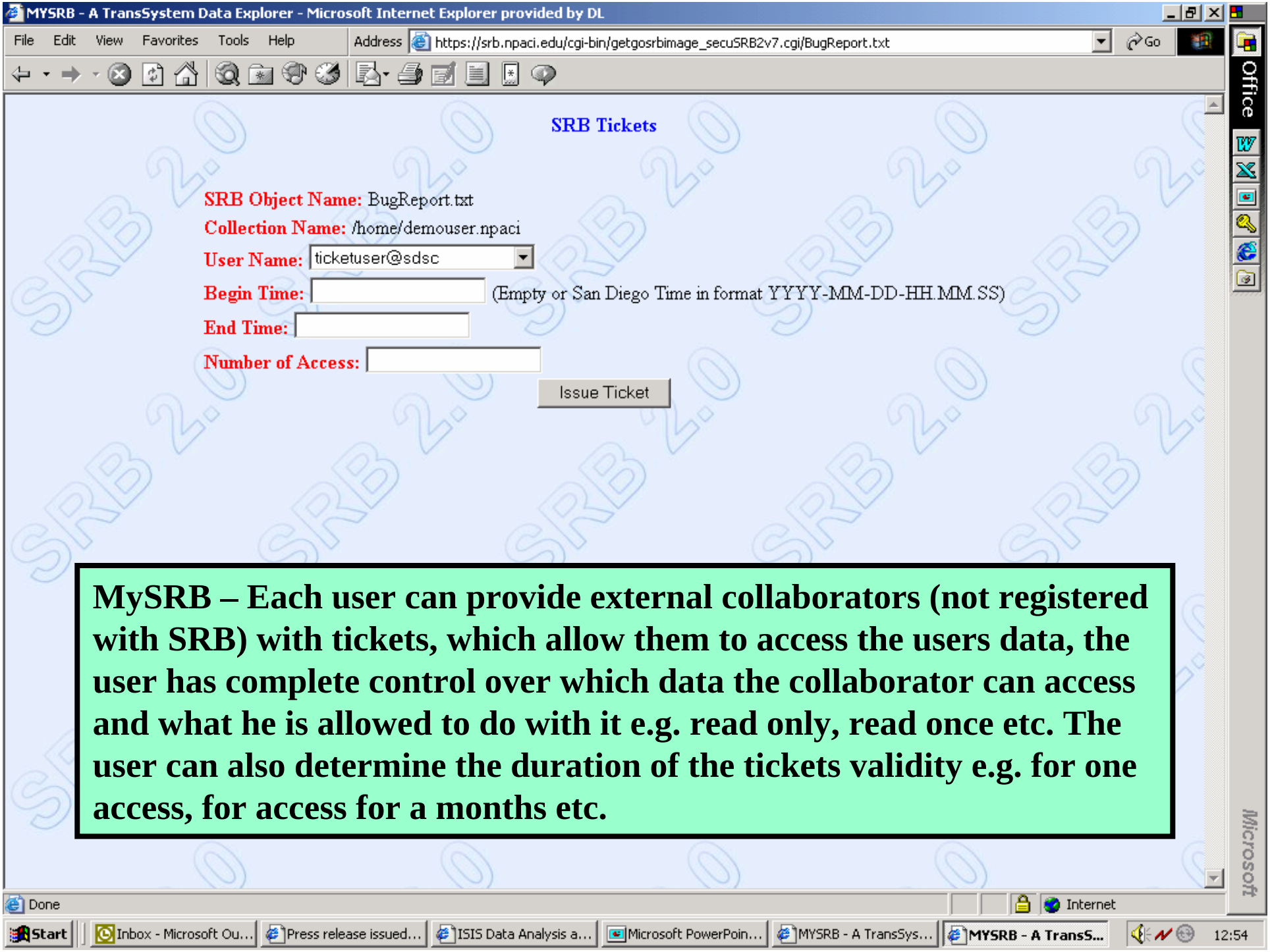
inQ – Allowing users to organise data into logical groupings independent of their physical location – see e.g. different resources in shown directory, one at Rutherford Appleton Laboratory and the other at Cambridge University.



MySRB – General view of users directory, functionalities accessible through buttons at the top or pull down menu next to each file. Again the files are in various different physical locations, but appear in one logical directory for easy usage.



MySRB – Ingestion of files into SRB from local machine, to a chosen resource within SRB.


MYSRB - A TransSystem Data Explorer - Microsoft Internet Explorer provided by DL
File Edit View Favorites Tools Help
Address https://srb.npaci.edu/cgi-bin/getgosrbimage_secuSRB2v7.cgi/BugReport.txt
Go
SRB Tickets
SRB Object Name: BugReport.txt
Collection Name: /home/demouser.npaci
User Name: ticketuser@sdsc
Begin Time:
(Empty or San Diego Time in format YYYY-MM-DD-HH.MM.SS)
End Time:
Number of Access:
Issue Ticket
MySRB – Each user can provide external collaborators (not registered with SRB) with tickets, which allow them to access the users data, the user has complete control over which data the collaborator can access and what he is allowed to do with it e.g. read only, read once etc. The user can also determine the duration of the tickets validity e.g. for one access, for access for a months etc.
Done
Internet
Start
Inbox - Microsoft Ou...
Press release issued...
ISIS Data Analysis a...
Microsoft PowerPoin...
MYSRB - A TransSys...
MYSRB - A TransS...
12:54

Usage Examples

Use Case 1 – Distributed Data Management / Archival Environment

CMS is one of the Large Hadron Collider (LHC) particle accelerator experiments at CERN. The CMS detector is one of the largest international scientific collaborations in history. As of February 2003 there were 2300 people working for CMS, 1940 of which were scientists and engineers. These people come from 159 institutes in 36 countries, spanning Europe, Asia, the Americas and Australasia. We are providing some of the data management and transfer backbone of the project with an SRB service.

Currently:

18 sites from Russia to the US

70 Physical resources

1000000 Files

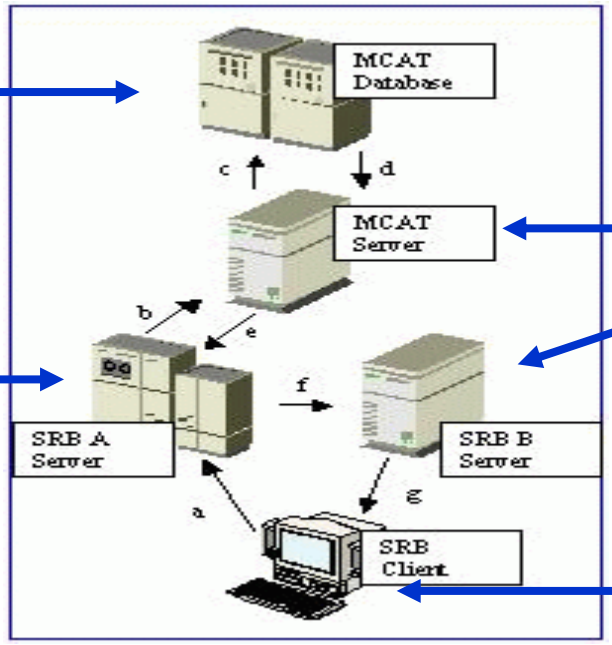
50TB of Data

Use Case 1 - Distributed Data Management / Archival Environment

Combining the storage capabilities of the Tape Stores around the world with the easy access and integration of the Storage Resource Broker.

Professional Databases for Central MCAT

Atlas Data Store accessed through SRB



Application Server to Access SRB Resource Groups and the Central MCAT

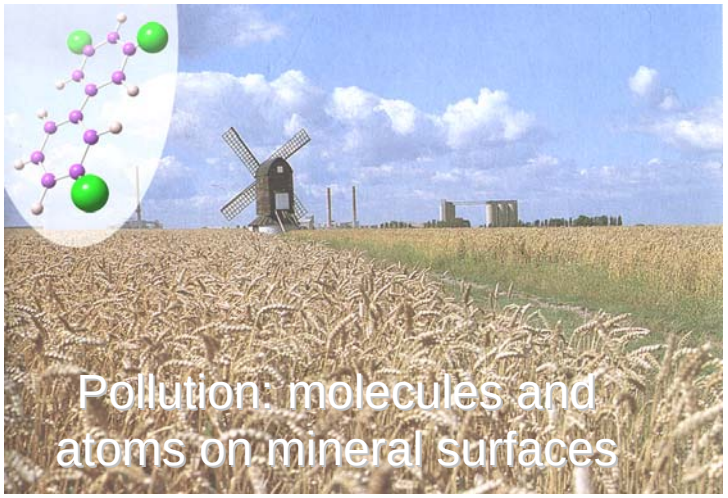
Transparent Access for users from their Desktop

Use Case 2 – Distributed Computing

e-Minerals: Environment from the Molecular Level

Modelling the atomistic processes involved in environmental issues

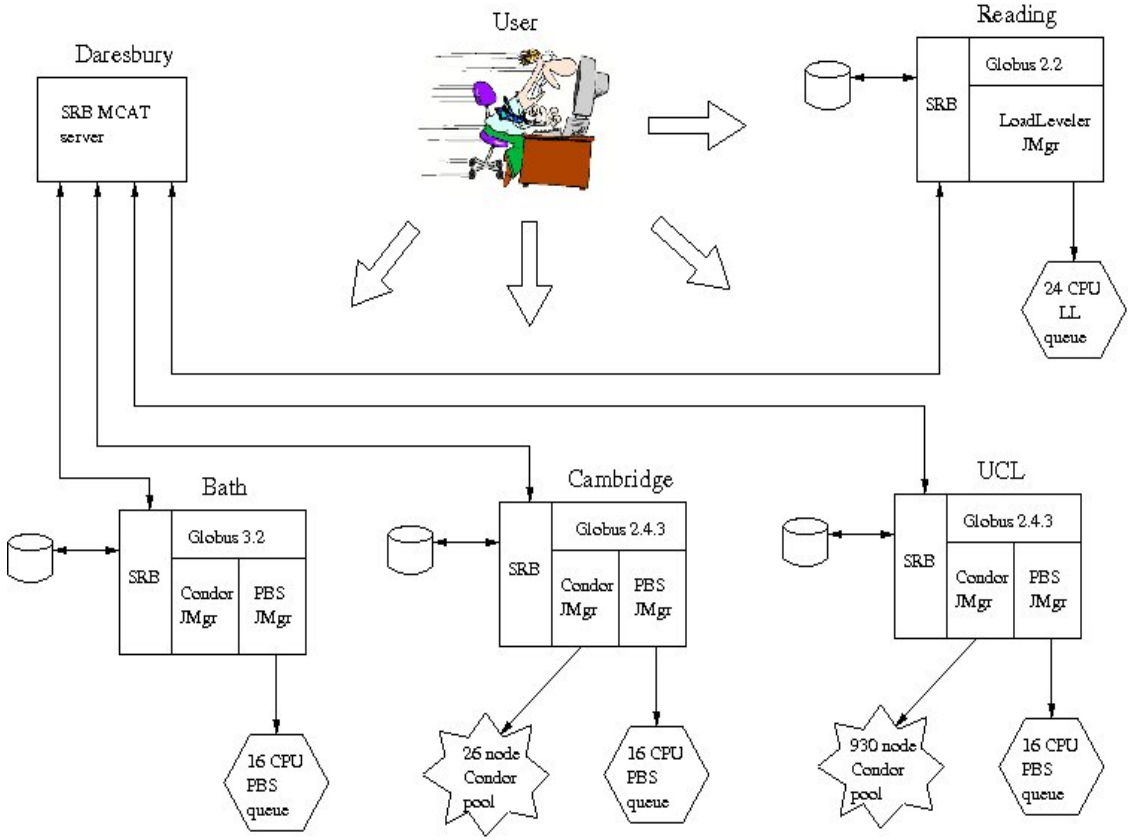
Radioactive waste disposal



Crystal growth and scale inhibition

http://www.eminerals.org

Use Case 2 – Distributed Computing (2)



e-Minerals Mini Grid

NGS

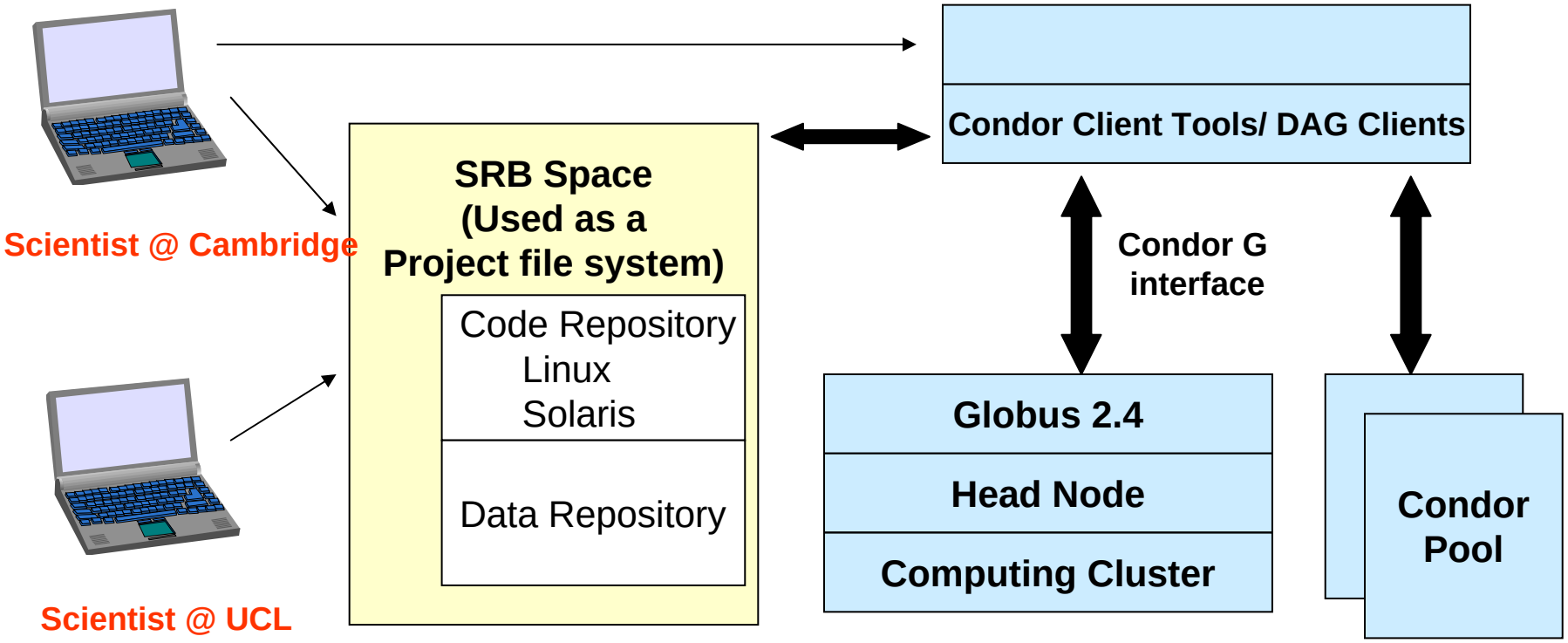
HPCx

CSAR

…..

UK Grid

Use of SRB on Eminerals



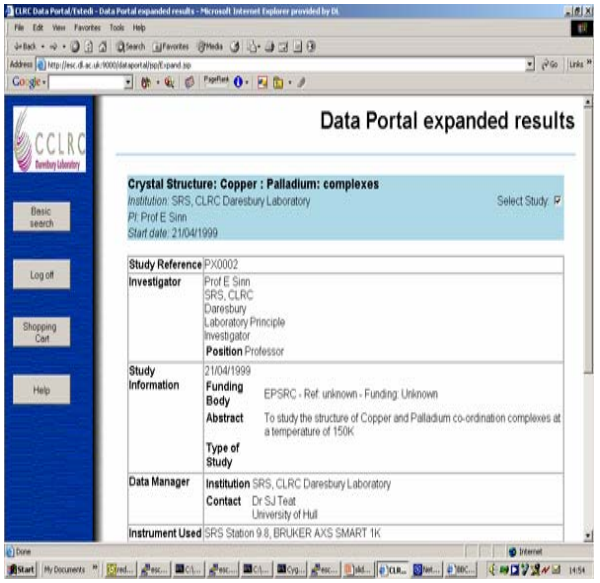
Use Case 2 – Distributed Computing (3)

- **Application are held in SRB – Directory for application, version + platform, as are standard set up scripts.**
- **User insures correct input data is available in SRB.**
- **User defines simulation(s – easily up to 1000) with Condor DaGMan Script incl. choosing system type – Pure Condor or Resource accessible via Globus e.g. with PBS scheduler and submits it from his desktop.**
- **The Job script is send to Condor Job Manager and submitted directly or via Globus to relevant system. The application code and input data are retrieved via the local SRB access. Job is started.**
- **The resulting data for every job is transferred into the user specified directory with in SRB through the local access. Job finished.**
- **Results are available through the usual SRB interfaces for the user to look at or further work on.**

The CCLRC DataPortal

DataPortal – One stop shop to search for and access data from different organisations on heterogeneous systems in a uniform way. Allows parallel querying of various resources, offers personal permanent workspace to work with the data. The system is based on a web services architecture, connects well with other services and offers a high level of security.

http://www.e-science.clrc.ac.uk/web/projects/dataportal

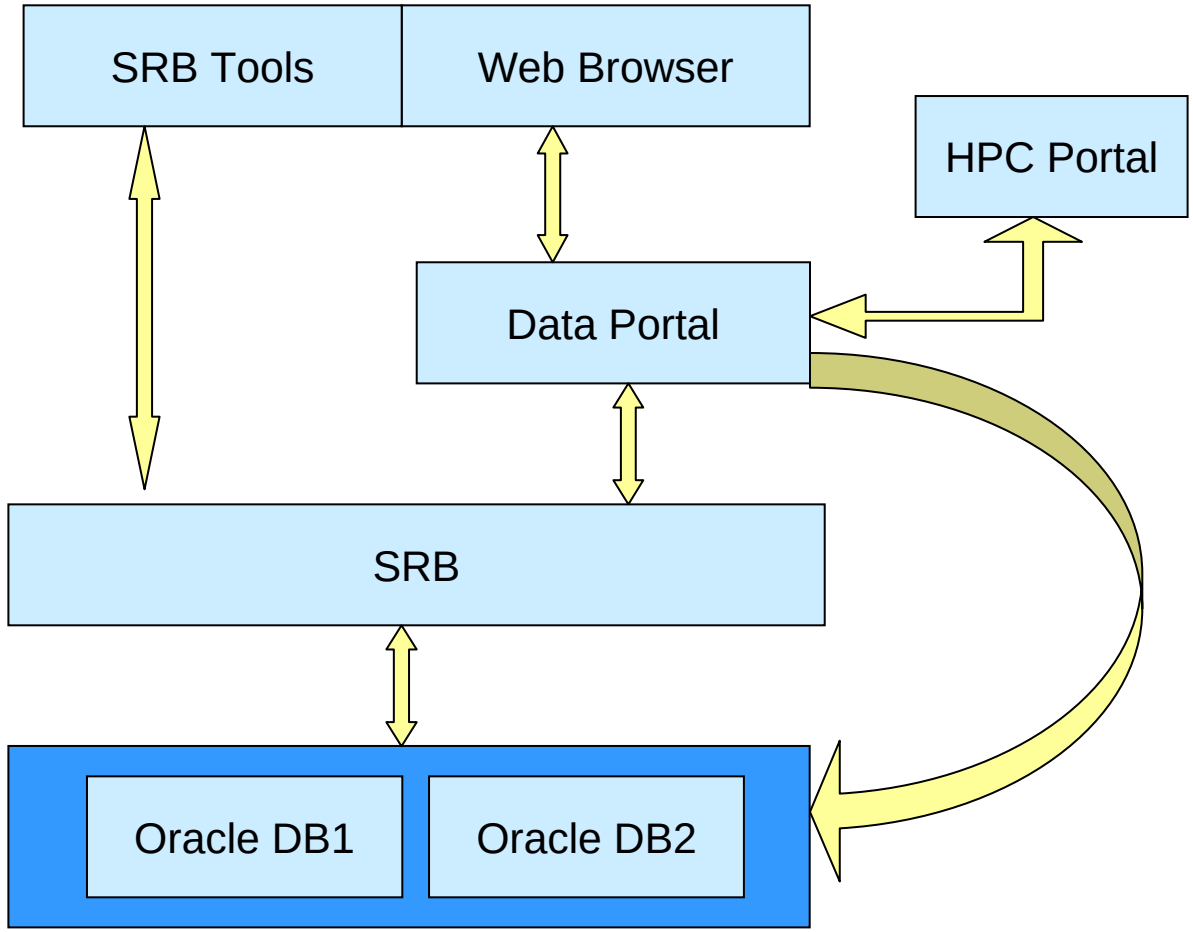


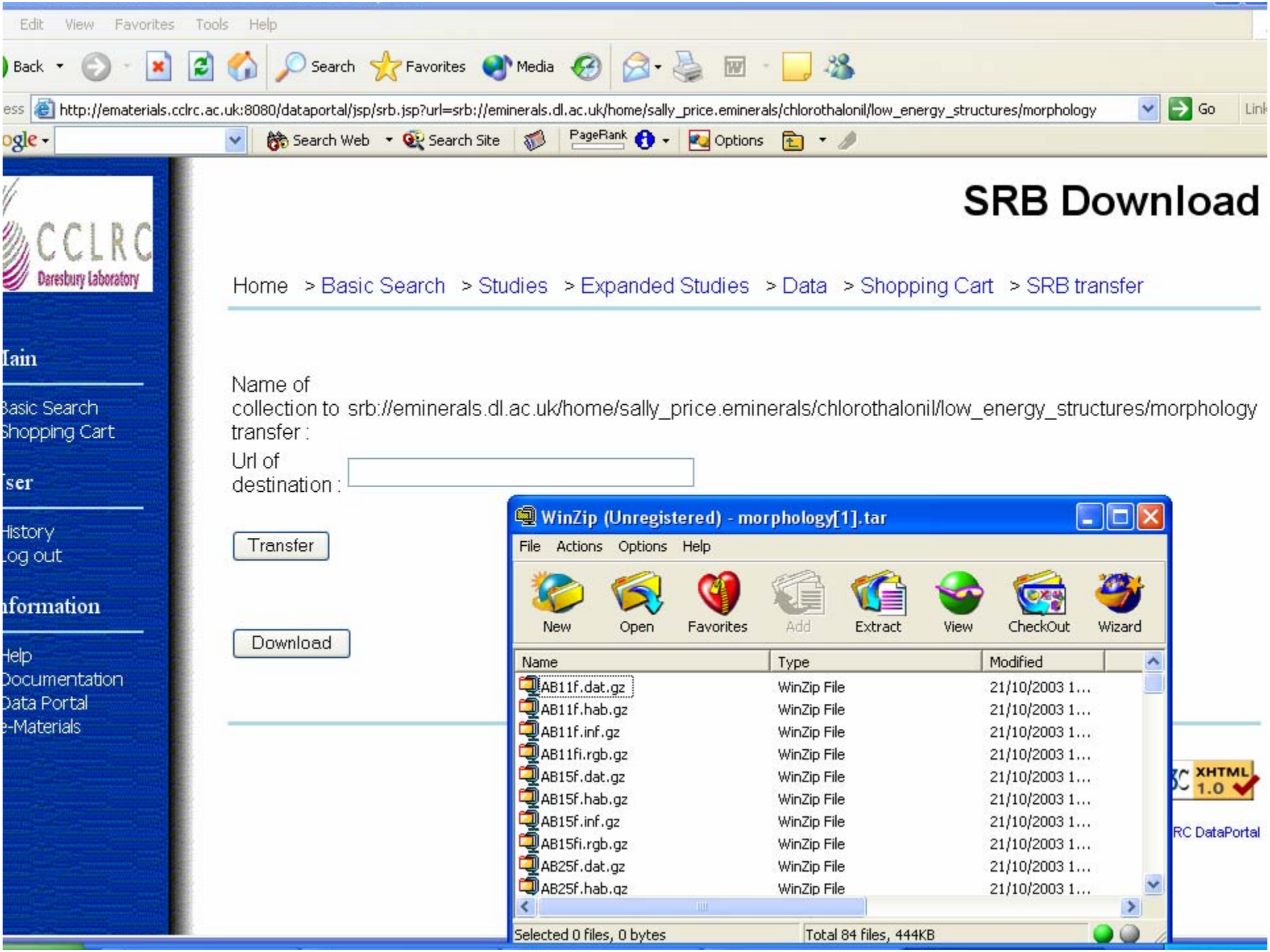
eMinerals integration with Data Portal

Files maybe inserted and retrieved using SRB tools

Also files may be accessed via the data portal if user Data Portal is also given access privileges

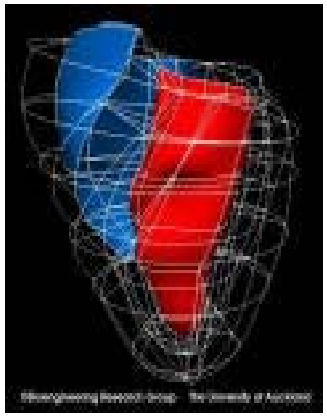
The Data portal may utilize different meta data to locate files



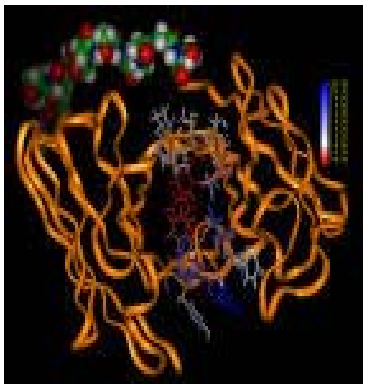

Edit View Favorites Tools Help
Back Search Favorites Media
http://ematerials.cclrc.ac.uk:8080/dataportal/jsp/srb.jsp?url=srb://eminerals.dl.ac.uk/home/sally_price.eminerals/chlorothalonil/low_energy_structures/morphology
Go
Search Web Search Site PageRank Options
CCLRC
Daresbury Laboratory
SRB Download
Home > Basic Search > Studies > Expanded Studies > Data > Shopping Cart > SRB transfer
Main
Basic Search
Shopping Cart
User
History
Log out
Information
Help
Documentation
Data Portal
e-Materials
Name of collection to transfer :
srb://eminerals.dl.ac.uk/home/sally_price.eminerals/chlorothalonil/low_energy_structures/morphology
Url of destination :
Transfer
Download
WinZip (Unregistered) - morphology[1].tar
File Actions Options Help
New Open Favorites Add Extract View CheckOut Wizard
Name Type Modified
AB11f.dat.gz WinZip File 21/10/2003 1...
AB11f.hab.gz WinZip File 21/10/2003 1...
AB11f.inf.gz WinZip File 21/10/2003 1...
AB11fi.rgb.gz WinZip File 21/10/2003 1...
AB15f.dat.gz WinZip File 21/10/2003 1...
AB15f.hab.gz WinZip File 21/10/2003 1...
AB15f.inf.gz WinZip File 21/10/2003 1...
AB15fi.rgb.gz WinZip File 21/10/2003 1...
AB25f.dat.gz WinZip File 21/10/2003 1...
AB25f.hab.gz WinZip File 21/10/2003 1...
Selected 0 files, 0 bytes
Total 84 files, 444KB
XHTML 1.0
RC DataPortal

Use Case 3 – Global Collaboration
Integrative Biology

An integrated approach to the giant killers..

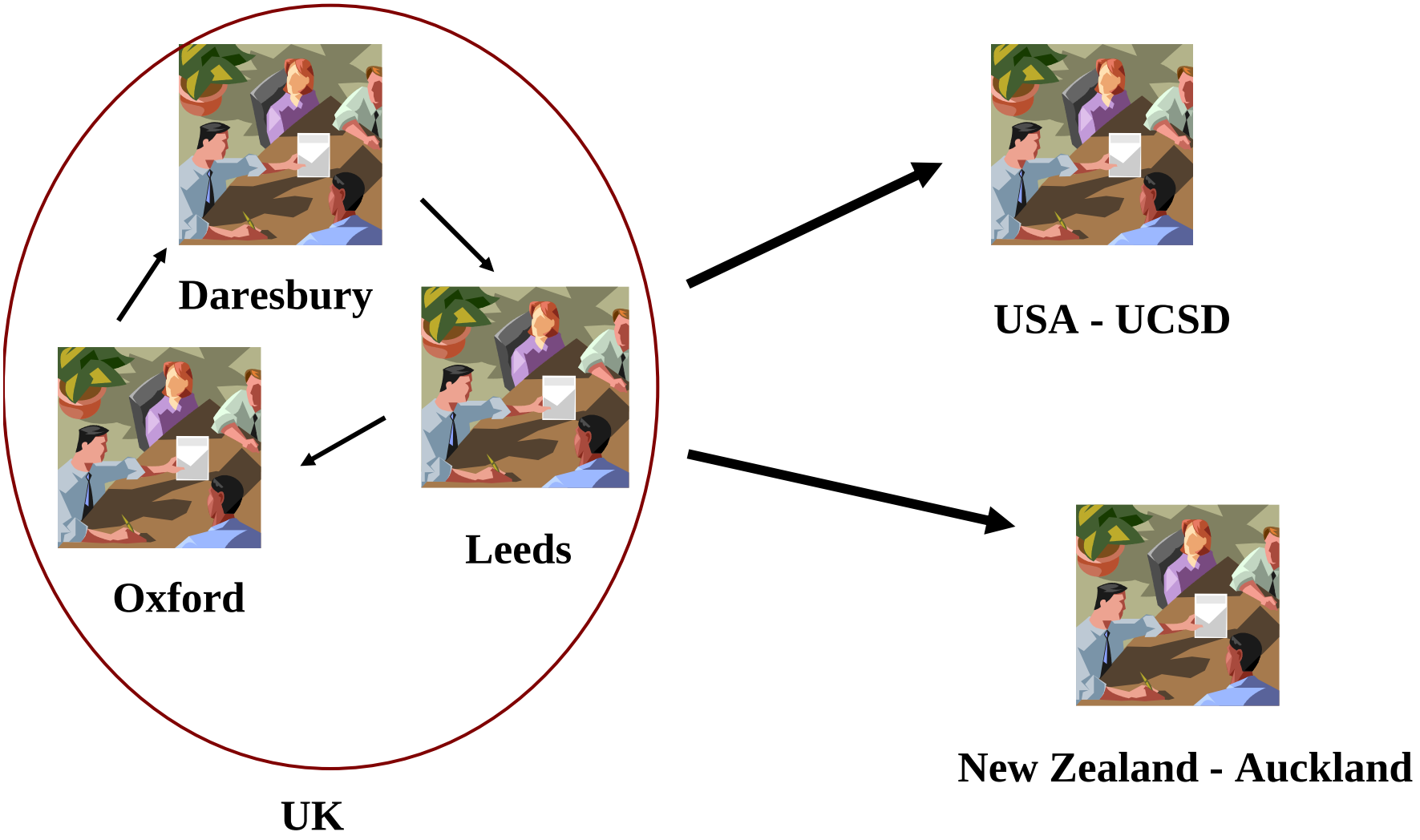


The Integrative Biology Grid will ultimately lead to better drugs for heart disease and cancer, two of the UK's biggest killers. A robust and fault-tolerant Grid infrastructure for biomedical science is the proposed aim of this recently announced EPSRC-funded e-Science project. Such a system will allow biomedical researchers to work seamlessly with distributed resources such as high-performance computers, databases and visualisation tools to develop complex models of how these killer diseases develop.



http://www.integrativebiology.ox.ac.uk/

Today's working environment (3)

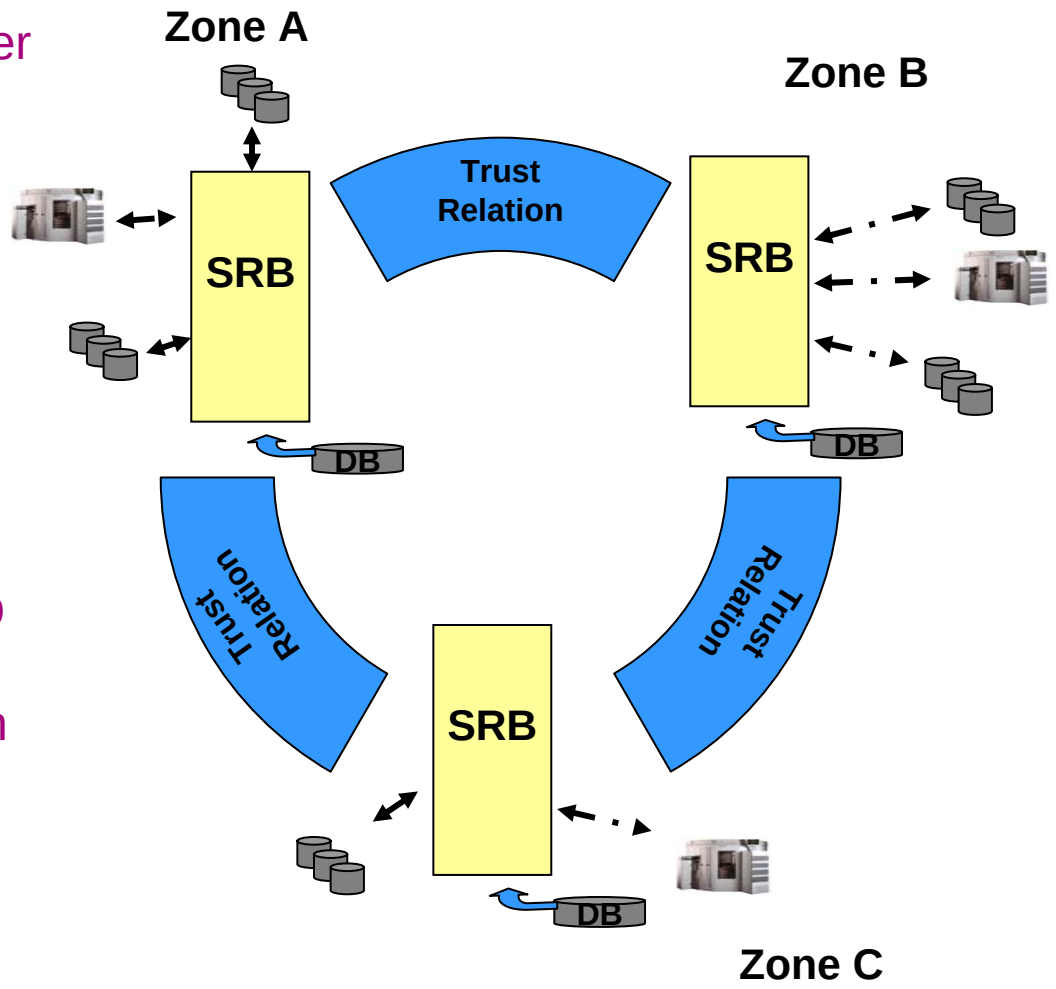


SRB Federation

Enables an SRB Network to recognize the presence of another SRB Network and be able to interact with it.

Every Zone is similar to a previous SRB deployment and is an independent administrative domain.

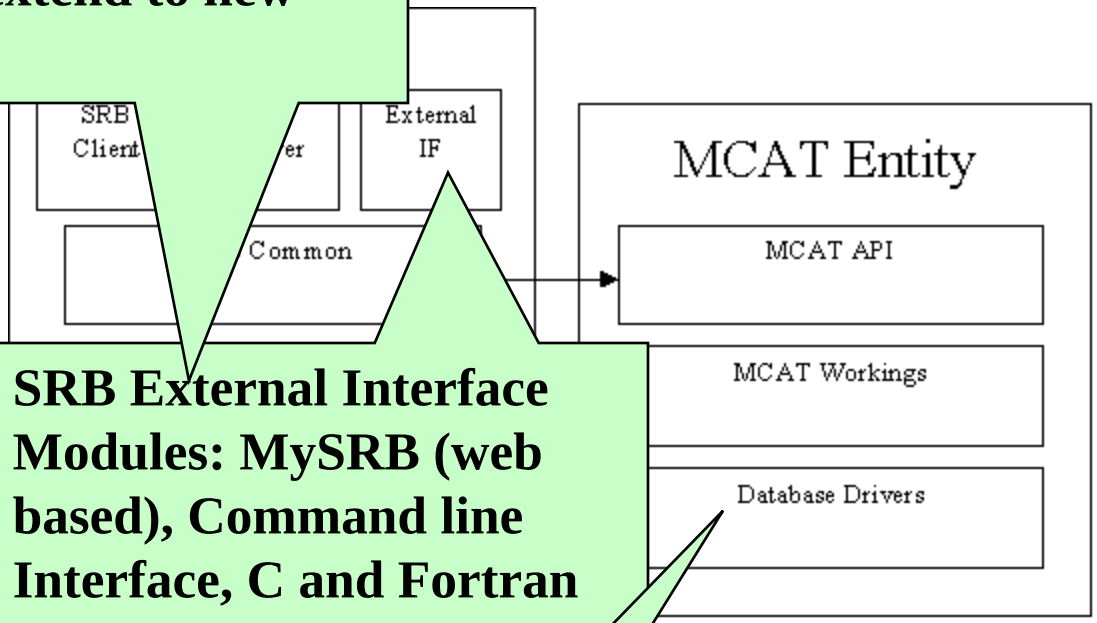
Based upon the trust relationship between the zones, user information and other information are exchanged for cross organizational collaboration



SRB – Behind the Scenes

roker – Internal Architecture

Devise Interface Modules to wide range of platforms – easy to extend to new systems



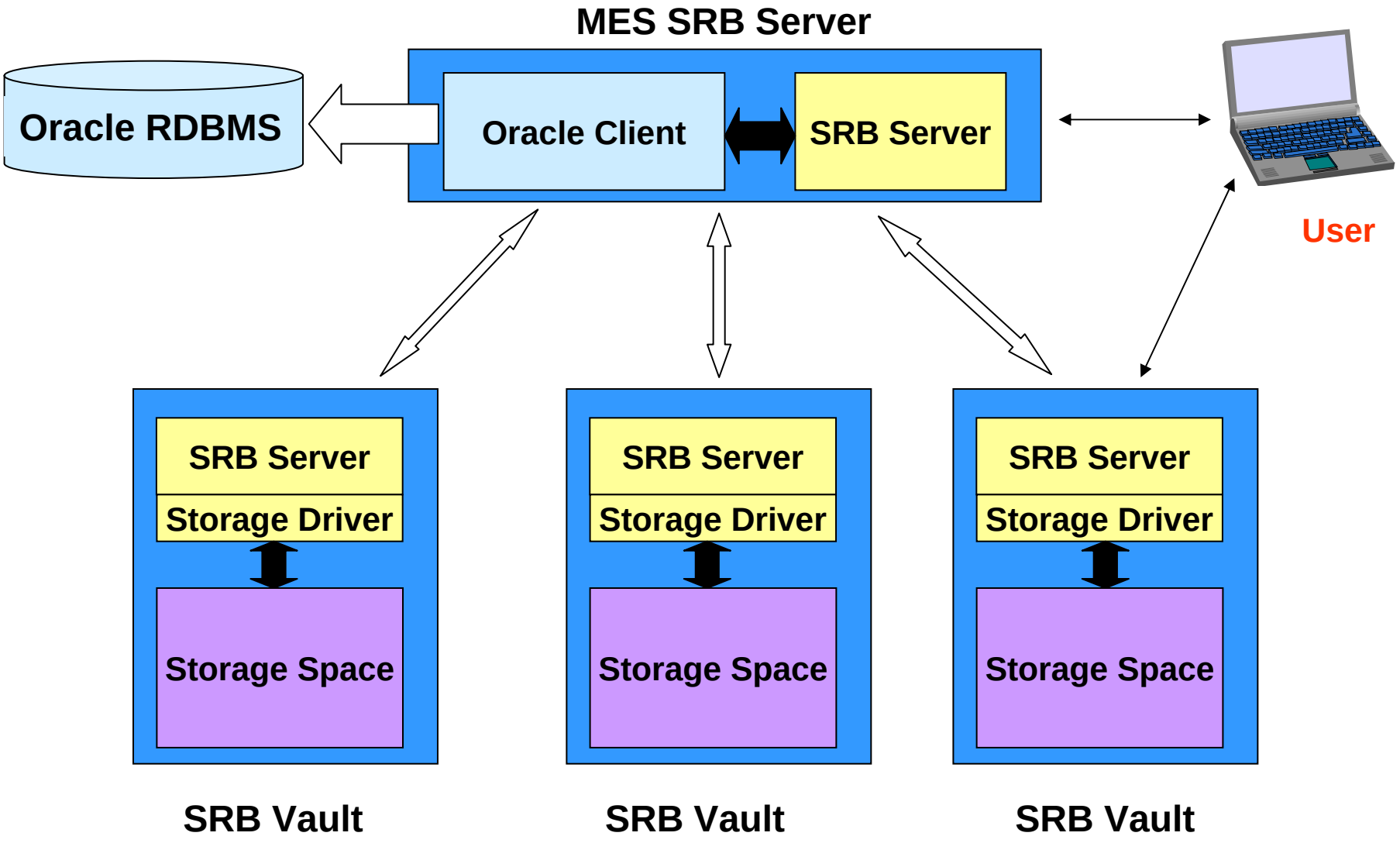
SRB External Interface Modules: MySRB (web based), Command line Interface, C and Fortran

MCAT manages links between logical to physical data location, replica and versioning. MCAT can be run on a variety of Relational Databases.

UNIX, LINUX, Databases

rb/mySRB/mySRB.html

SRB Physical Structure



Behind the scenes SRB provides many other functionalities in managing files and resources

- Fine Grained Access Control to files
- Meta Data Query and File Replication between resources
- Grouping of multiple physical resources into a logical resource.
- Direct Client Server parallel file transfers for performance improvements
- Supports grouping of multiple files into ‘containers’ which is then manageable for insertion and retrieval from Mass Storage systems.

Storage Resource Broker – Main Features

•SRB provides an uniform API that can be used to connect to heterogeneous resources that may be distributed and access data sets that may be replicated.

•SRB allows users to manage data storage and replication across the wide range of physical storage system types and locations, while still allowing having a single, stable, access point to the data. SRB has two major components, the core SRB, which interfaces with the storage devices, and the MCAT, which holds the metadata elements.

•Many different platforms and authentication methods are supported by the modular design, and a web service interface is available.

•The system provides interfaces for the ingestion of data and associated metadata; management of replication and data movement; searching the metadata for discovery; and the retrieval of the data itself. Metadata held to support these interfaces includes the physical and logical details of the data held and its replicas, user information, and security rights and access control.

Thank you for you attention.

Any questions??

Contact details

http://www.e-science.clrc.ac.uk

k.kleese@dl.ac.uk