ARCHER Overview

October 2008

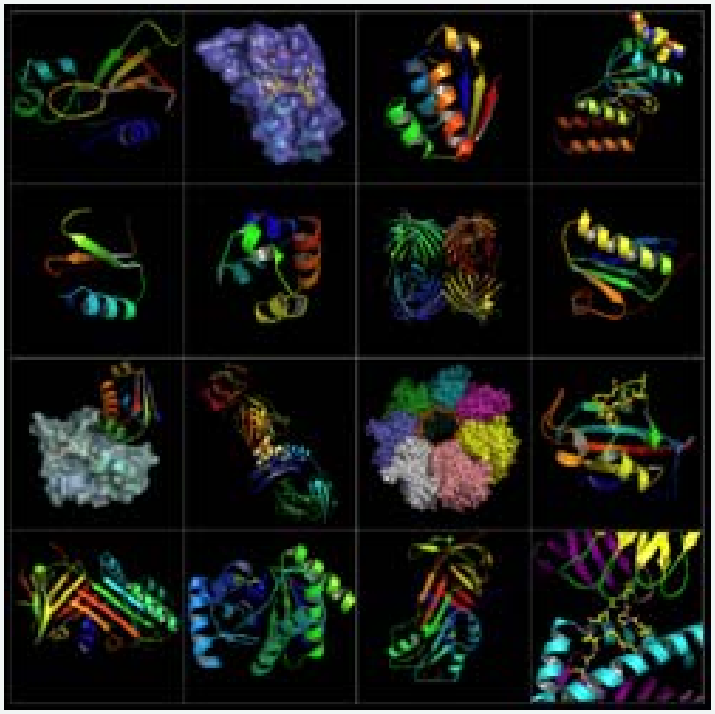
MONASH University | JCU James Cook University | The University Of Queensland Australia

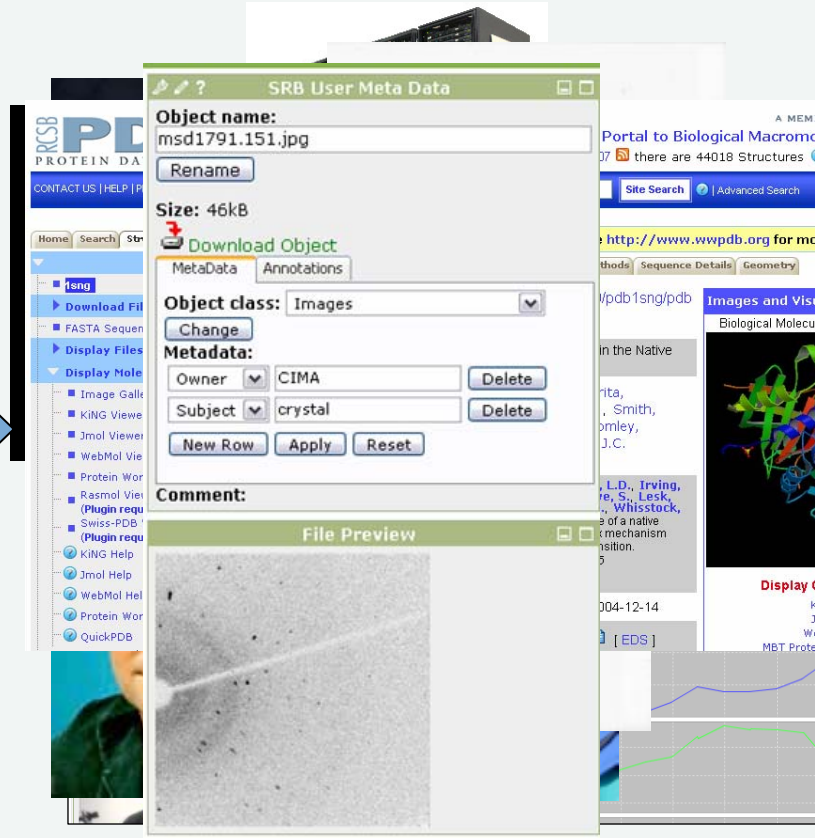
archer.edu.au

e-Research Challenges

- Acquiring data from instruments
- Storing and managing large quantities of data
- Processing large quantities of data
- Sharing research resources and work spaces between institutions
- Publishing large datasets and related research artifacts
- Searching and discovering

Research process





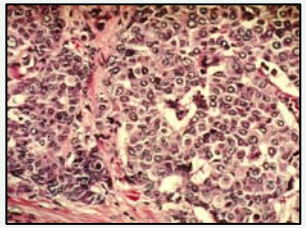
Crystal exposed to X-rays & diffraction pattern detected

Results [illegible] repositories

ARCHER - Australian ResearCh Enabling enviRonment

- Building generic research data management infrastructure:
  - ARCHER Research Repository
  - Distributed Integrated Multi-Sensor & Instrument Middleware – concurrent data capture and an
  - Scientific Dataset Manager (Web)
  - Scientific Dataset Manager (Desktop Client)
  - Metadata Editing Tool
  - Analysis Workflow Automation Tool
  - Collaborative and Adaptable Research Portal Development Tool
- Work on Shibboleth enhancements and security requirements with the AAF
- Completed some customised deployments
- Developed by Monash University, James Cook University, and University of Queensland
- Funded by DIISR/DEST, through the SII (Systemic Infrastructure Initiative)
- ARCHER will be completed by September 2008

Instruments



Acquire

Manual

ARCHER

Building generic tools for a secure, seamless, and collaborative e-Research space

- Dataset Acquisition
- Dataset Management (Web)
- Dataset Management (Desktop)
- Collaborative Workspaces
- Workflow Automation
- Metadata Management

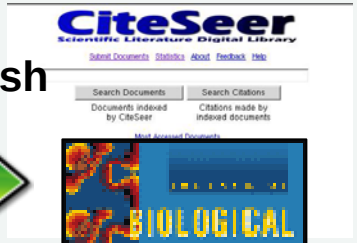
Research Repositories

Computational Grids

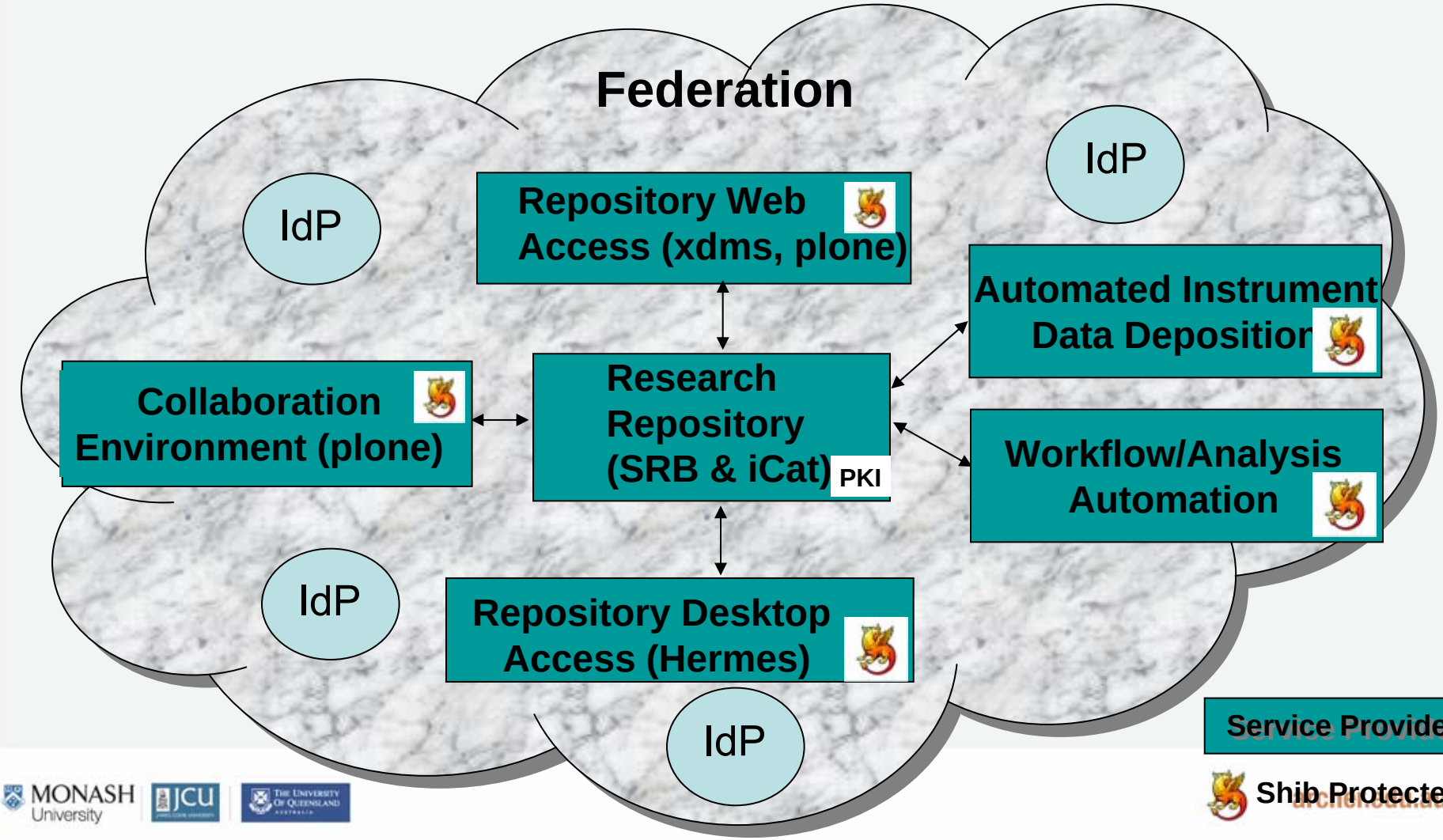
Publication Repositories

arrow
australian research repositories online to the world

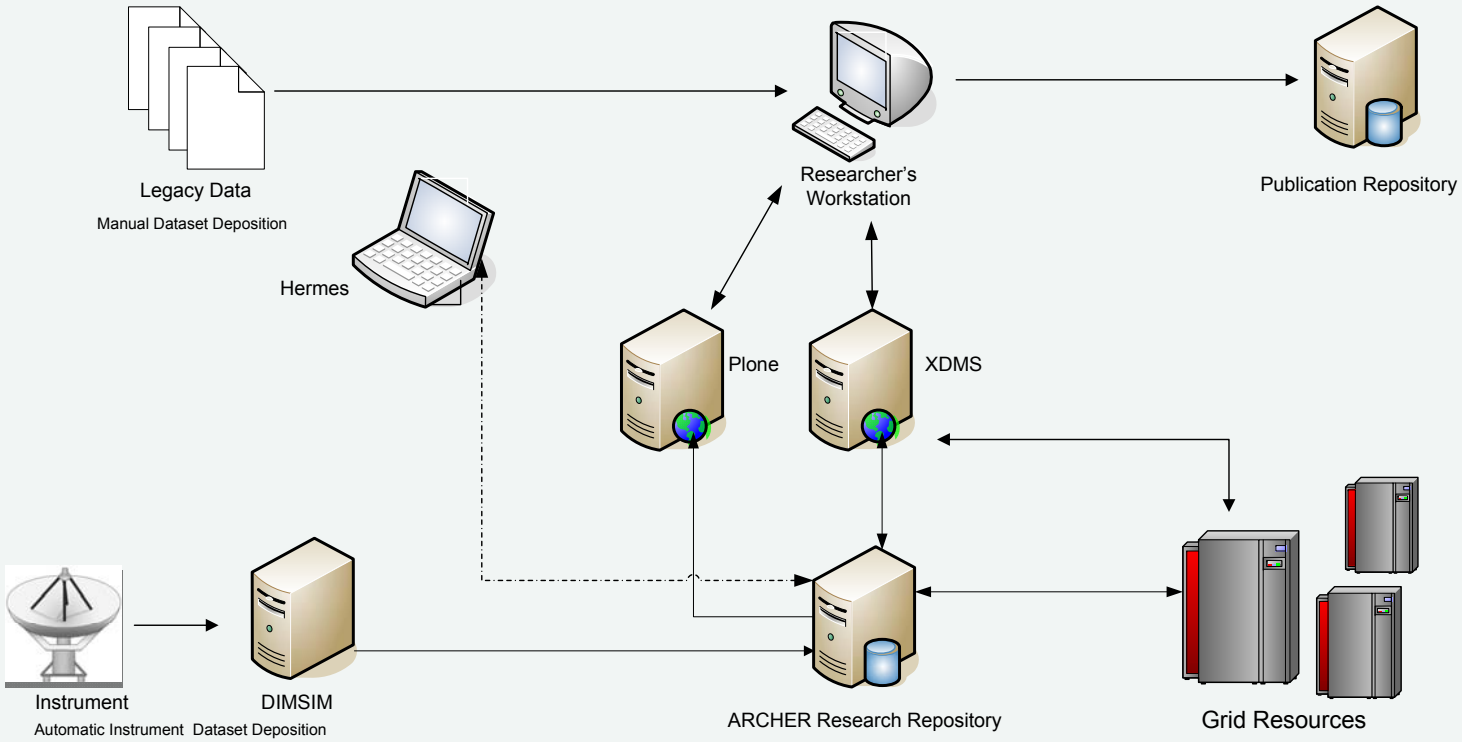
Publish



ARCHER: Data-centric Model



An Example ARCHER Deployment



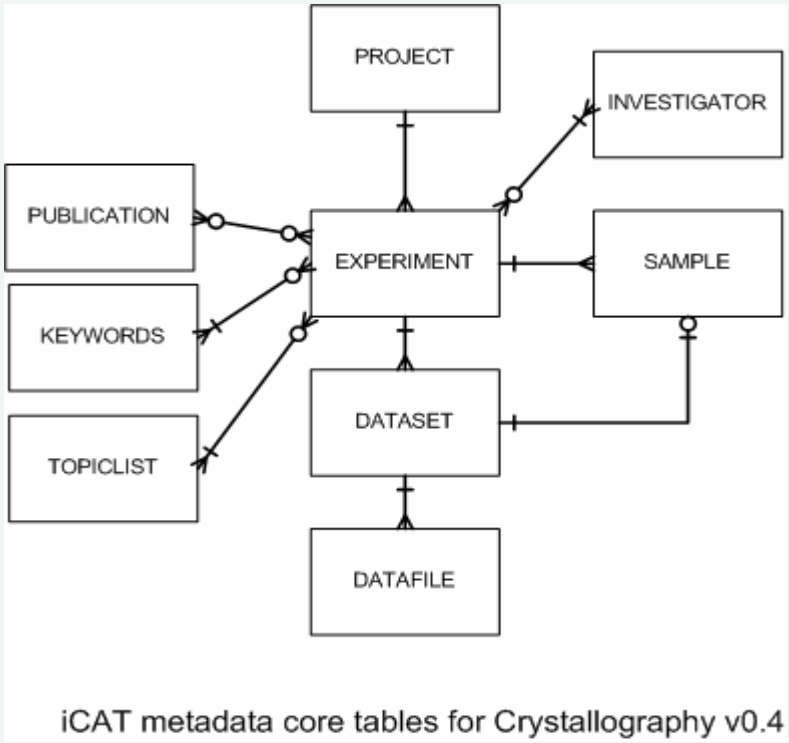
ARCHER Research Repository

A place for Researchers to store their research data

- Easily Accessible
  - Federated access - aligns with the AAF
  - Research data can be accessed by web, desktop, or standard file access protocols (e.g. GridFTP and SRB)
- Capable of managing large datasets
  - Built on SRB
- Rich metadata
  - Core metadata based on CCLRC's* Scientific Metadata Model
  - Flexible metadata available for samples, datasets, and datafiles
- Secure

* Now the Science and Technology Facilities Council

Simplified CCLRC Scientific Metadata Model



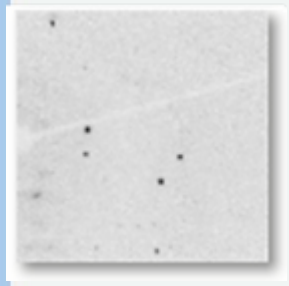
iCAT metadata core tables for Crystallography v0.4

Distributed **I**ntegrated **M**ulti-**S**ensor & **I**nstrument **M**iddleware (DIMSIM)

Concurrent data capture & analysis

- Allows multiple sensors to be easily integrated
- Enables instruments to be more easily accessible over a network
- Automatically deposits instrument datasets into a designated research repository
- Easily accessible telemetry
- Enables concurrent analysis

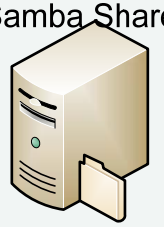
DIMSIM for Crystallographers



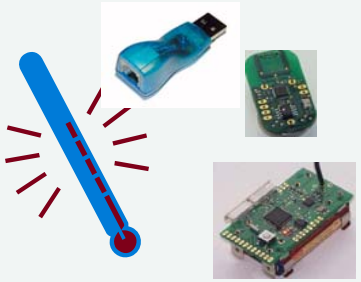
OSC Images

Diffractometer

Rigaku Control
Samba Share

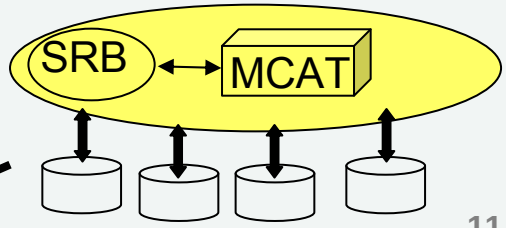


Images



Sensors (lab environment etc.)

Common Instrument
Middleware Architecture
(CIMA)



Disk/Tape Storage
(multiple locations)

Useful stuff!

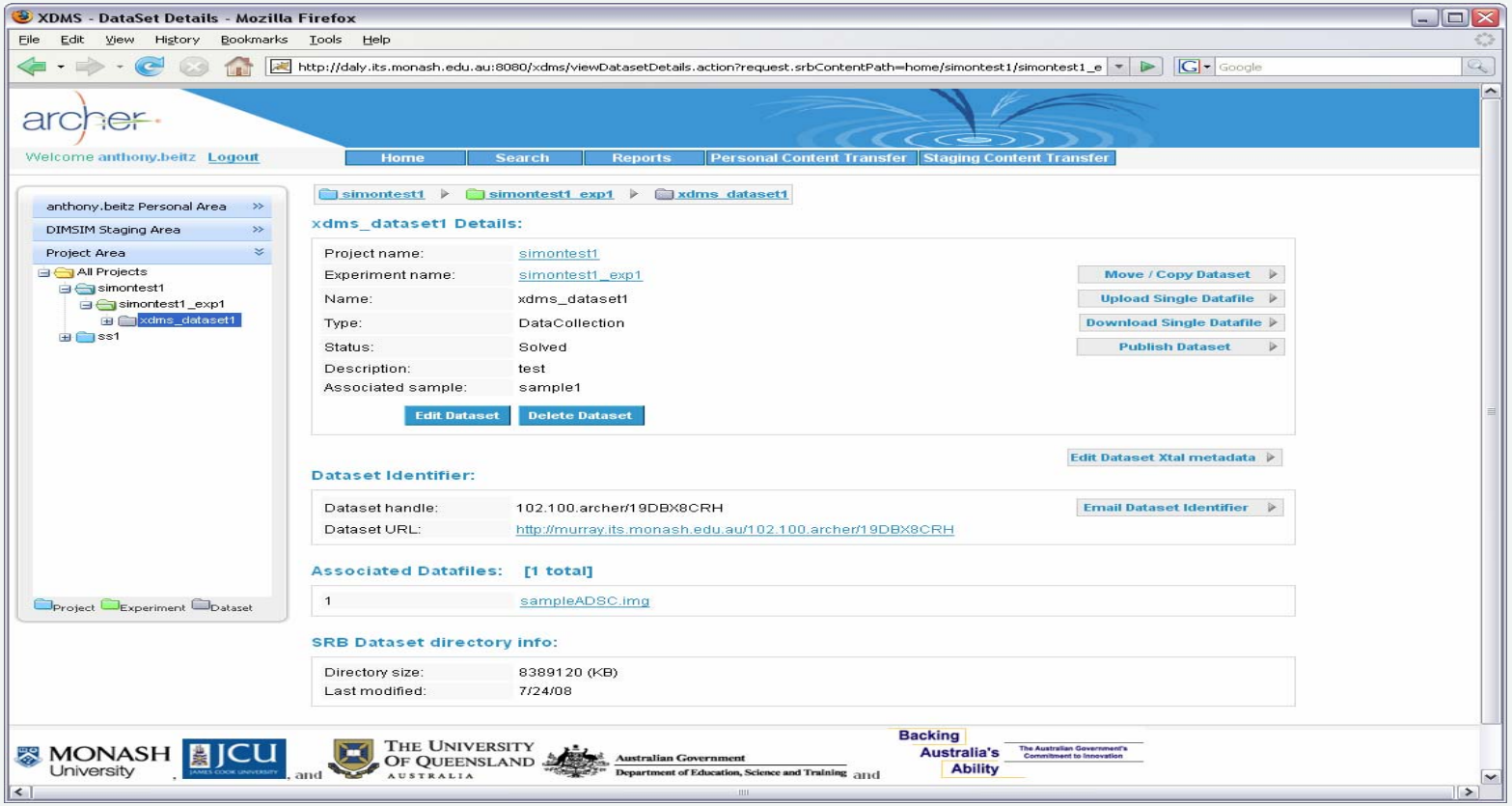
DIMSIM – Example Telemetry

XDMS: Scientific Dataset Manager (Web)

A web tool for Researchers to manage and curate their research data

- Formalised research data management
  - Directory structure follows CCLRC's Scientific Metadata Model
  - Suitable for dataset collection/analysis/publication
  - Create/Read/Update/Delete support
- Powerful search capabilities
- Automatic metadata extraction from research datafiles
- Rich metadata editing capabilities (via MDE)
- Secure and accessible
  - Federated access
    - Aligns with the AAF (Australian Access Federation)
  - Protected by Shibboleth
- Utilises Handles (persistent identifiers) for external links
- Dataset export to Fedora

XDMS


XDMS - DataSet Details - Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://daly.its.monash.edu.au:8080/xdms/viewDatasetDetails.action?request.srbContentPath=home/simontest1/simontest1_e
archer
Welcome anthony.beitz Logout
Home
Search
Reports
Personal Content Transfer
Staging Content Transfer
anthony.beitz Personal Area
DIMSIM Staging Area
Project Area
All Projects
simontest1
simontest1_exp1
xdms_dataset1
ss1
Project Experiment Dataset
simontest1 simontest1_exp1 xdms_dataset1
xdms_dataset1 Details:
Project name: simontest1
Experiment name: simontest1_exp1
Name: xdms_dataset1
Type: DataCollection
Status: Solved
Description: test
Associated sample: sample1
Edit Dataset
Delete Dataset
Move / Copy Dataset
Upload Single Datafile
Download Single Datafile
Publish Dataset
Edit Dataset Xtal metadata
Dataset Identifier:
Dataset handle: 102.100.archer/19DBX8CRH
Dataset URL: http://murray.its.monash.edu.au/102.100.archer/19DBX8CRH
Email Dataset Identifier
Associated Datafiles: [1 total]
1 sampleADSC.img
SRB Dataset directory info:
Directory size: 8389120 (KB)
Last modified: 7/24/08
MONASH University
JCU
and
The University Of Queensland AUSTRALIA
Australian Government
Department of Education, Science and Training
and
Backing Australia's Ability

Metadata Editing Tool (MDE)

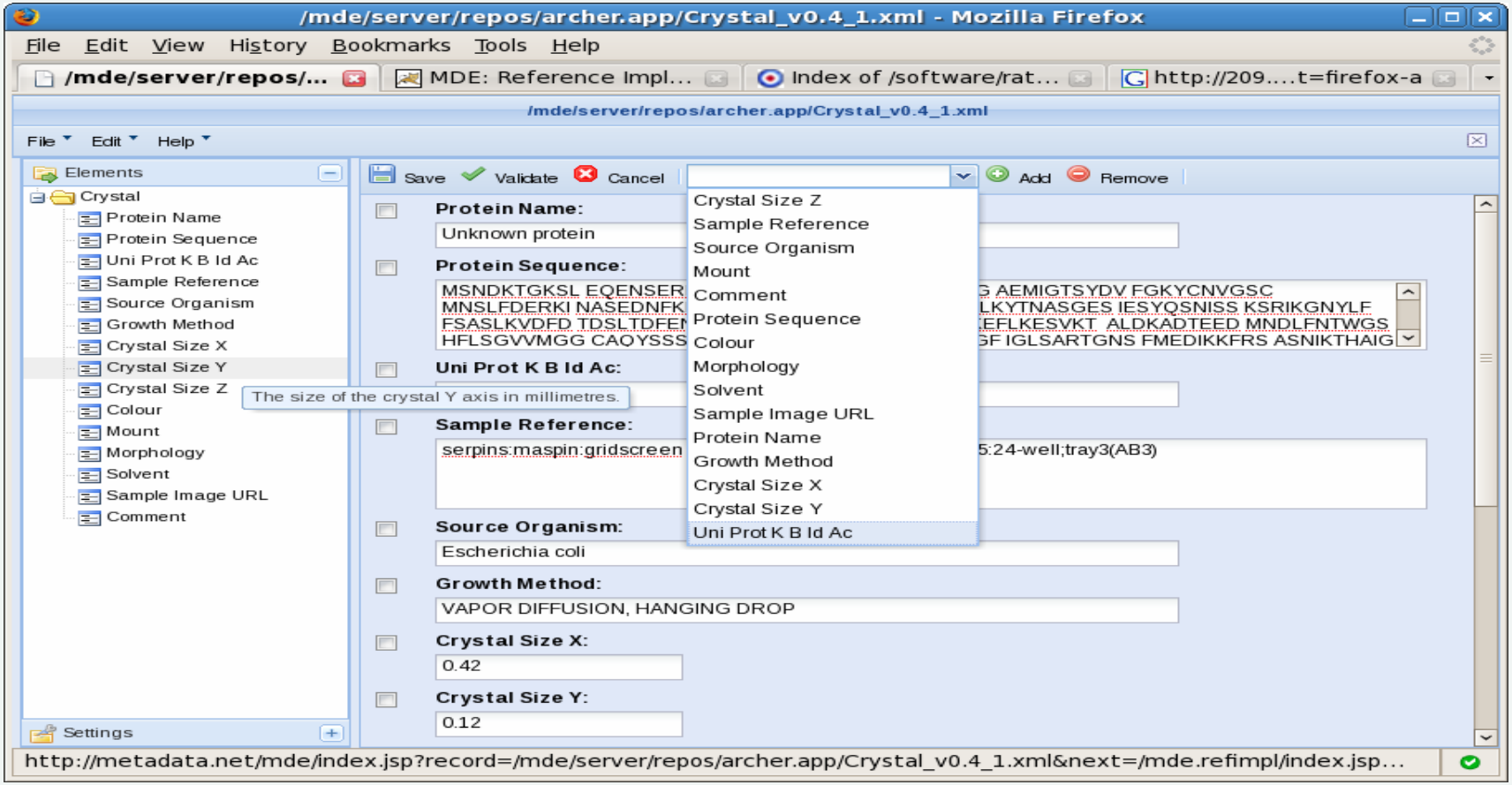
Schema driven metadata editing for e-Research

- The key innovation of MDE is that it is a schema-driven editor.
- MDE uses the schema to build a Web 2.0 form layout for the metadata. The layout includes the following:
  - Form elements for displaying the existing metadata elements, with type-specific input controls for entering the values. These include such things as number and date validation, and pull-downs for controlled lists.
  - Element descriptions available as hover-text.
  - Controls for creating and deleting elements based on what the schema allows and requires.

When the user decides to save the metadata record, it undergoes complete validation against the schema. The validation process checks that:

- the elements in the record are all defined in the schema and present in the correct number,
- the values of the elements satisfy any type restrictions defined by the schemas; e.g. elements defined as integers should consist of digits with an optional leading sign,
- schema-specific constraints on the record and individual elements are all satisfied.

Metadata Editing Tool

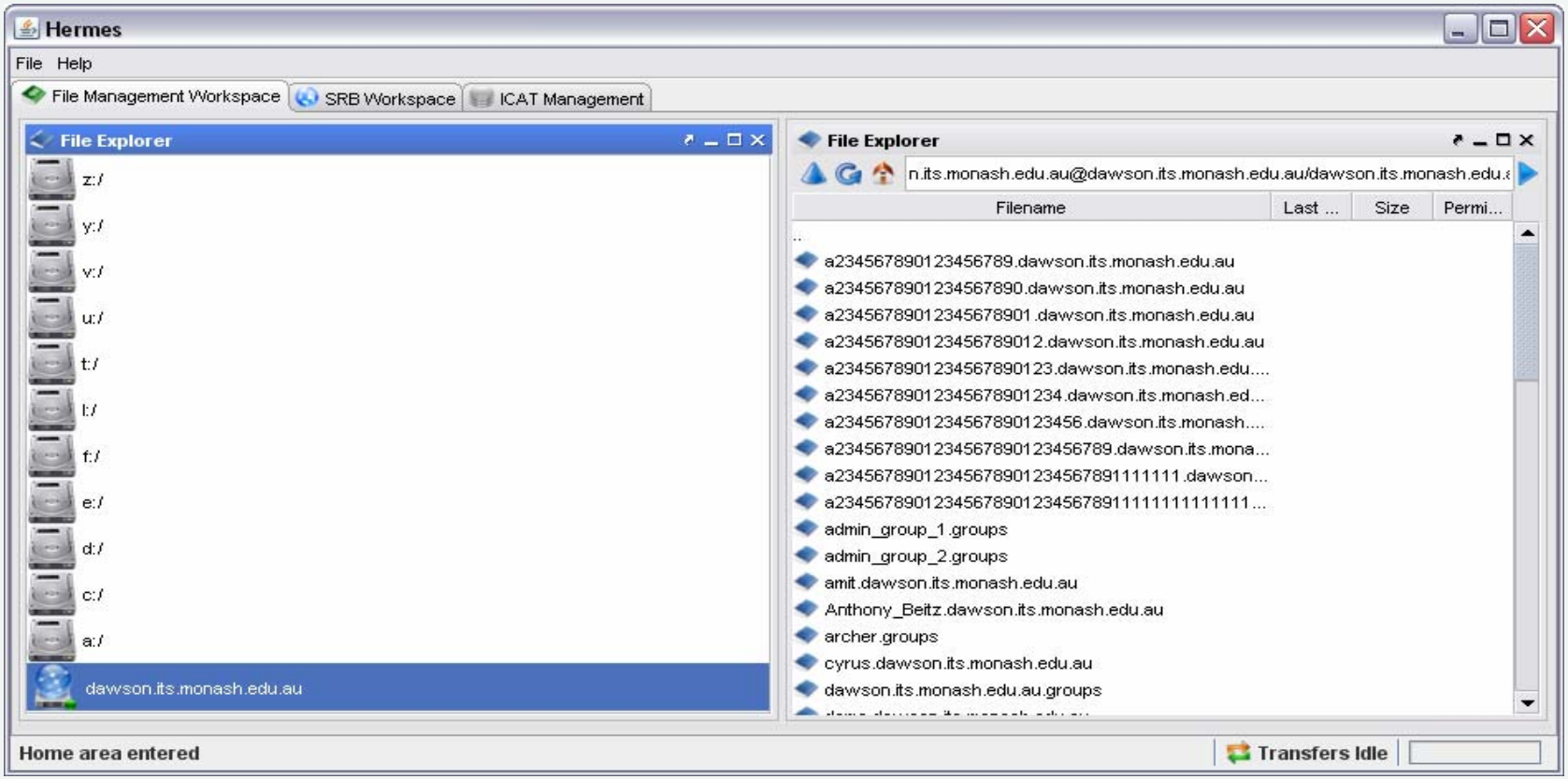


Hermes: Scientific Dataset Manager (Desktop Client)

A desktop tool for Researchers to transfer/manage their research data

- Doesn't have timeout issues for large data transfers that web apps experience
- Platform-independent (written in Java)
- Federated access
  - Aligns with the AAF (Australian Access Federation)
  - Protected by Shibboleth and PKI technologies
- Dock-able file browser
- Supports many different types of file systems (gftp, srb,cifs etc.)
- Freedom to access the storage system of choice
- Supports plugins, which interface to the institutions metadata repository.
- Addition of customised views of metadata repositories

Hermes


Hermes
File Help
File Management Workspace
SRB Workspace
ICAT Management
File Explorer
z:/
y:/
v:/
u:/
t:/
l:/
f:/
e:/
d:/
c:/
a:/
dawson.its.monash.edu.au
File Explorer
n.its.monash.edu.au@dawson.its.monash.edu.au/dawson.its.monash.edu.
Filename
Last ...
Size
Permi...
a234567890123456789.dawson.its.monash.edu.au
a23456789012345678901.dawson.its.monash.edu.au
a2345678901234567890123.dawson.its.monash.edu....
a234567890123456789012345.dawson.its.monash.ed...
a23456789012345678901234.dawson.its.monash....
a2345678901234567890123456.dawson.its.mona...
a234567890123456789012345678911111111.dawson...
admin_group_1.groups
admin_group_2.groups
amit.dawson.its.monash.edu.au
Anthony_Beitz.dawson.its.monash.edu.au
archer.groups
cyrus.dawson.its.monash.edu.au
dawson.its.monash.edu.au.groups
Home area entered
Transfers Idle

Hydrant: Analysis Workflow Automation Tool

Streamlining Analysis

- Web based portal which sits on top of the core Kepler engine
- Easy for researchers to reproduce or modify an analysis
  - Analysis is described by a workflow
  - Workflow is in XML form and can be presented on the web visually
  - Workflow can be executed on a workflow engine from the web
  - Researchers can easily modify aspects of workflow from the web
  - Researchers can share their workflows
- Secure and accessible

Hydrant

Hydrant - Mozilla Firefox
File Edit View History Bookmarks Tools Help
Customize Links Free Hotmail Windows Marketplace Windows Media Windows GridSphere Portal XDMS Dev XDMS Closed Ticket R...
hydrant
the scientific workflow hub
Anthony Beitz
Dashboard Workflows Jobs Profile FAQ Log out
Hello World
Create Job | Download | Duplicate | Delete
The classic "Hello World" in Kepler. This workflow simply prints a constant value to a display window.
Hello World
SDF Director
Constant
Display
The classic "Hello World" in Kepler. This workflow simply prints a constant value to a display window.
Matt Jones, May 2006, NCEAS
Hello World job submit preview
count
1
The number of times to print the text
Text
"Hello World"
The text to print.
Hosted by:
JCU
django POWERED

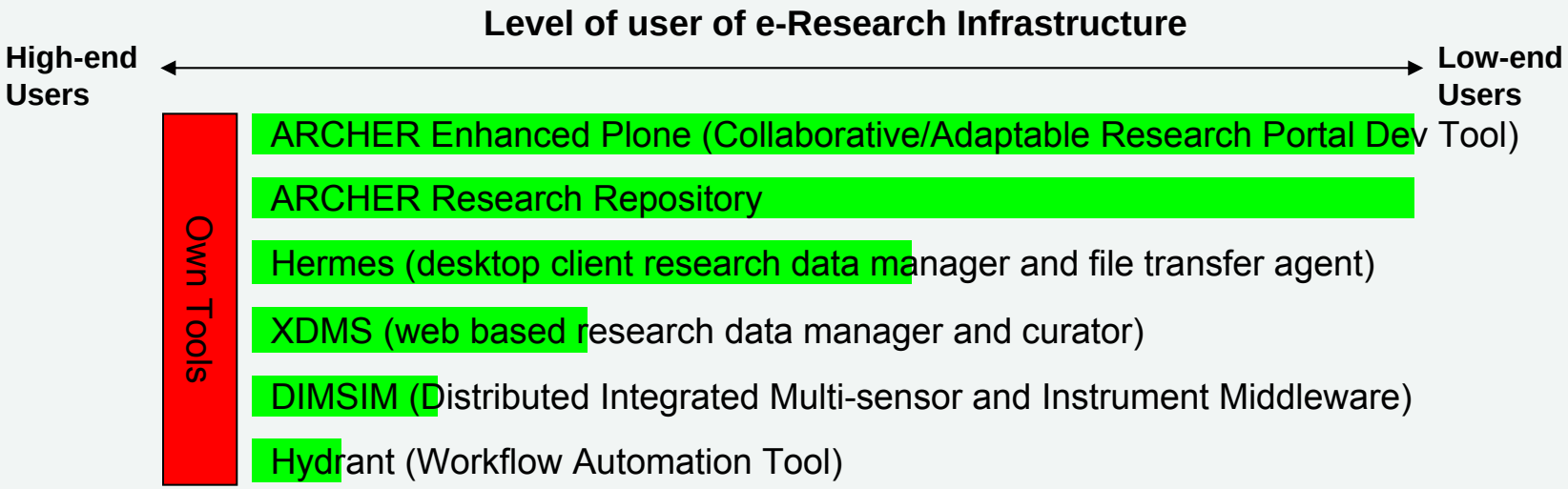
ARCHER Enhanced Plone: Collaborative and Adaptable Research Portal Development Tool

Bringing Researchers together

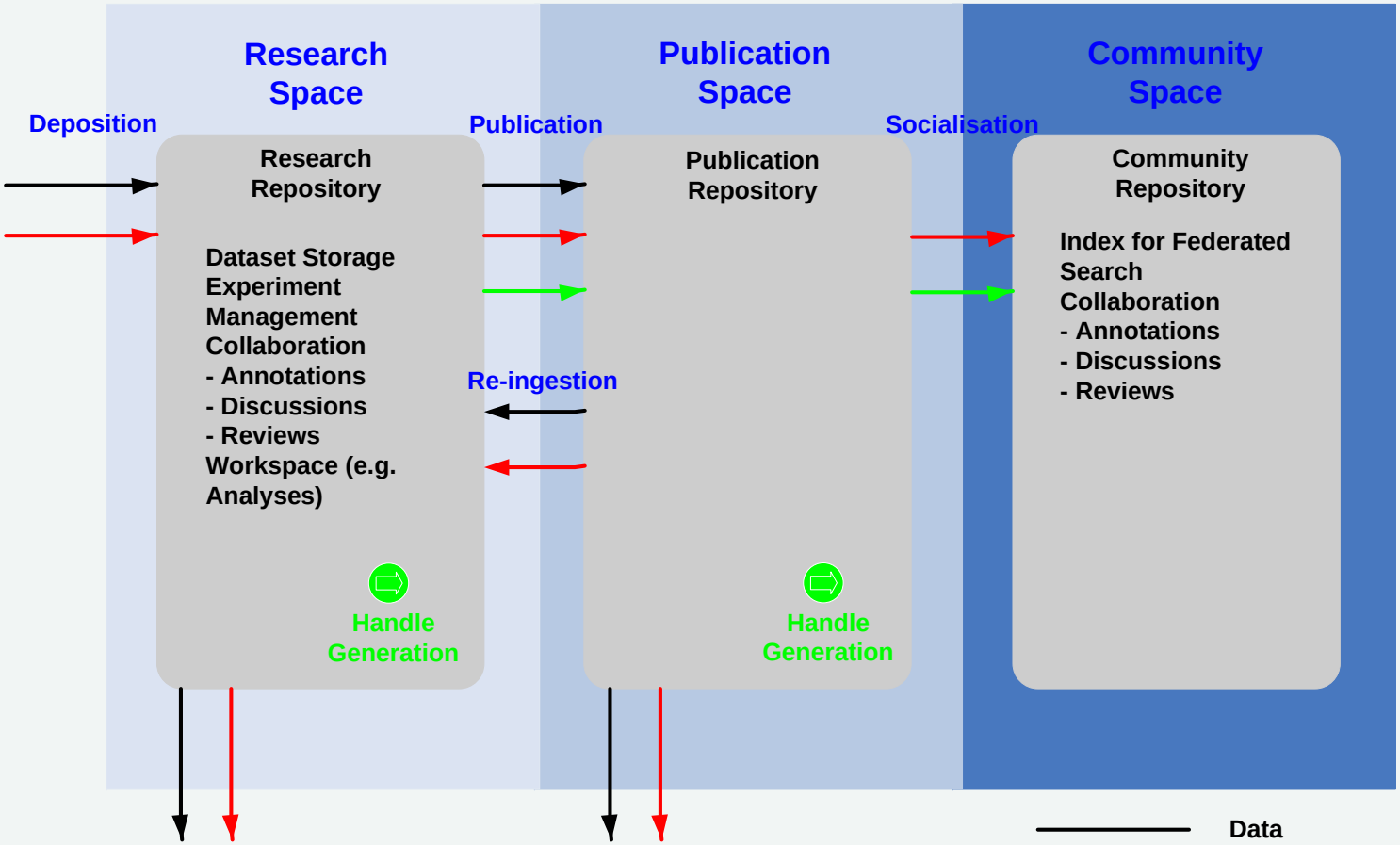
- Simplifies research portal development
  - Easy to author and manage own web content
- Enables sharing, management, and discussions of documents
- Built on Plone
  - Open source Content Management System (CMS)
- Powerful search capabilities
- Secure and accessible
  - Federated access - aligns with the AAF (Australian Access Federation)
  - Protected by Shibboleth
- Access to the ARCHER Research Repository

Plone – an Example Portal

ARCHER Expected Tool Usage



e-Research Repository Space



TARDIS

THE AUSTRALIAN REPOSITORIES FOR DIFFRACTION IMAGES
(Androulakis et al (2008) Acta Cryst. D) [PubMed]

Welcome

TARDIS is a multi-institutional collaborative venture that aims to facilitate the archiving and sharing of raw X-ray diffraction images (collectively known as a 'dataset') from the Australian protein crystallography community. Whereas the model coordinates and (less often) the structure factors (processed experimental data) are stored in the PDB the raw diffraction data is often not available. Addressing this shortcoming is the central aim of TARDIS. There are several reasons why we feel this is important (see about), which can be summarised as:

- The availability of raw data is extremely useful for the development of improved methods of image analysis and data processing.
- Fostering the archival of raw data at an institutional level is one the best ways of ensuring that this data is not lost (laboratory archives are typically volatile)

We have currently finished development on a suite of deposition tools that allow X-ray diffraction images to be deposited in an open source Fedora repository, and are about to provide discovery services to index the resulting data. Much of this site is under development, however, you can find some basic deposited data here.

For more information on the TARDIS project please contact Dr. Ashley Buckle.

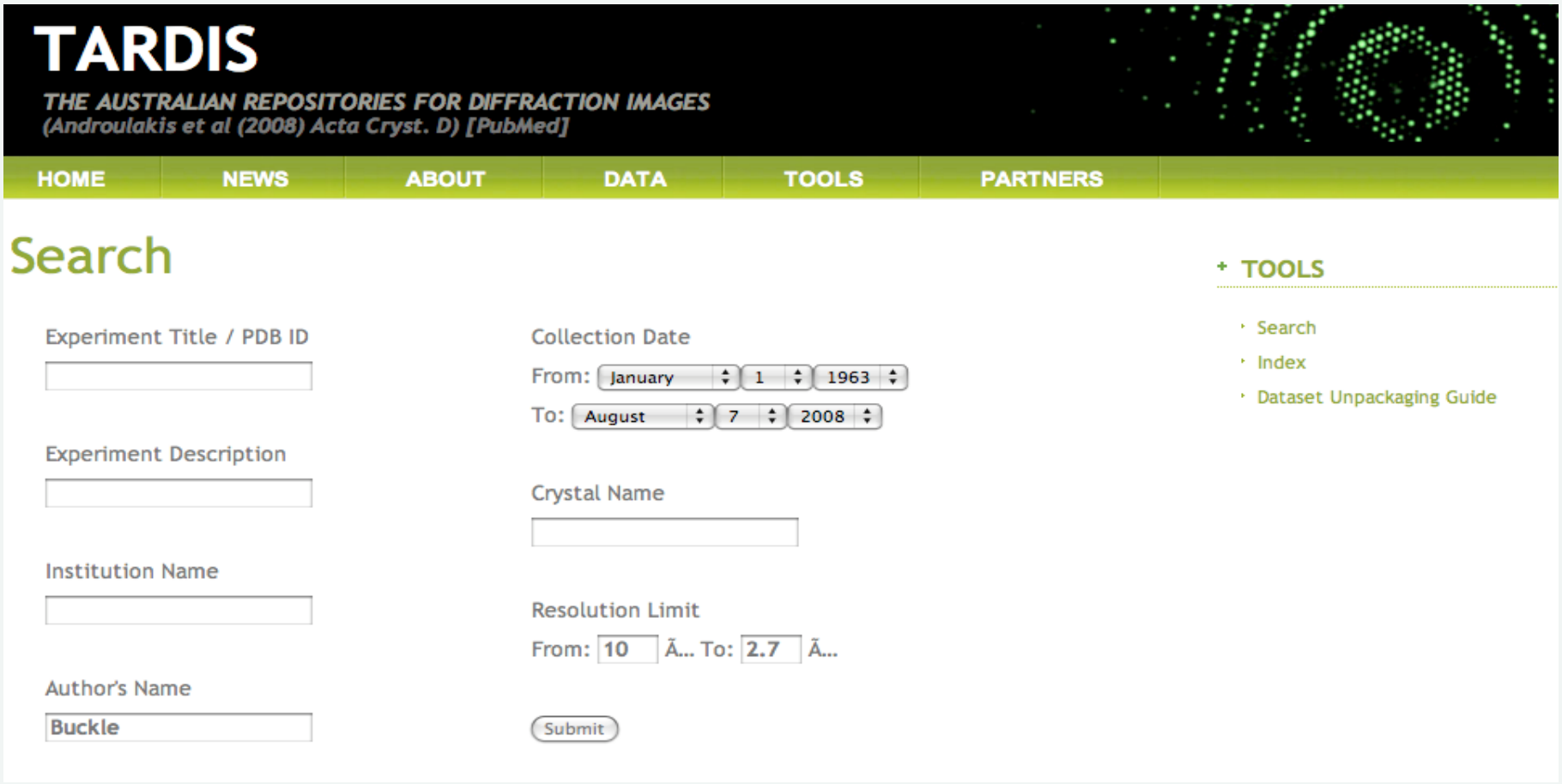
+ RECENT NEWS

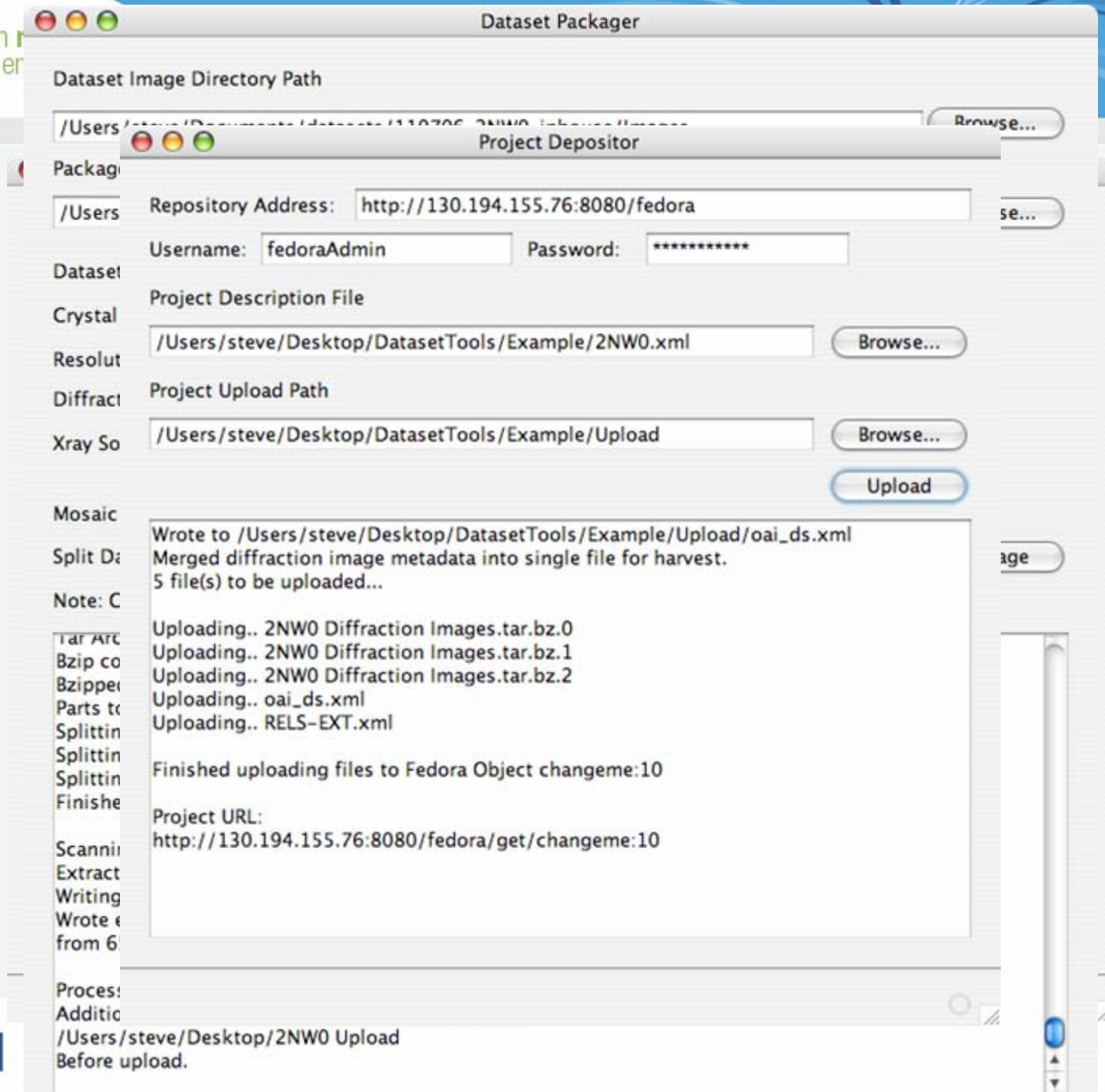
- TARDIS Beta Portal
- Acta Cryst. Paper Online
- Dataset Tools Released
- TARDIS Mentioned In Acta Cryst
- Site Revamp
- Deposition Tools Coming Soon!
- Site Update

PARTNERS

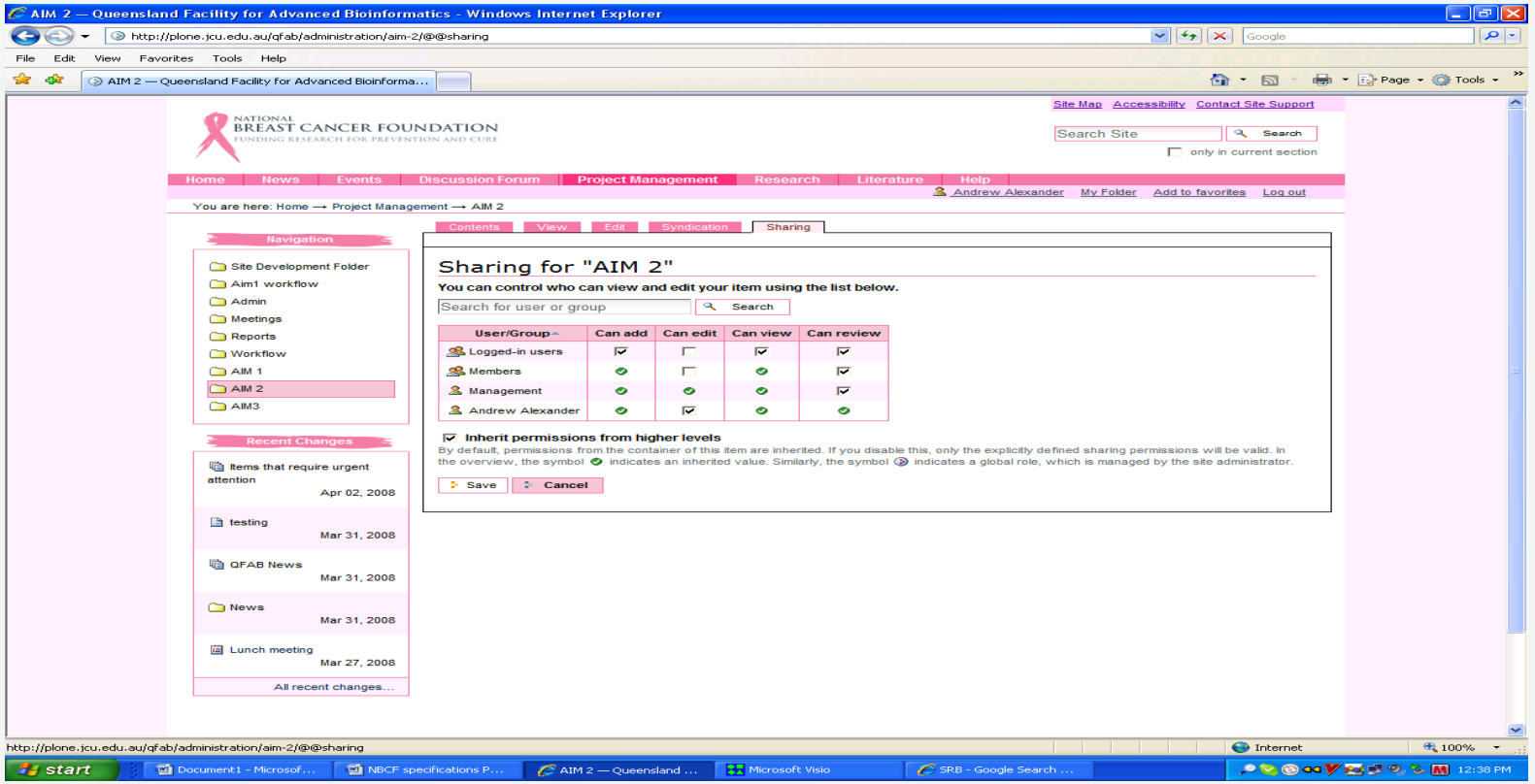
MONASH University
bio21 institute
The University of Melbourne
The University Of Queensland AUSTRALIA
arrow archer australian research repositories online to the world
St Vincent's Health
IMB

Discipline Specific Federated Search



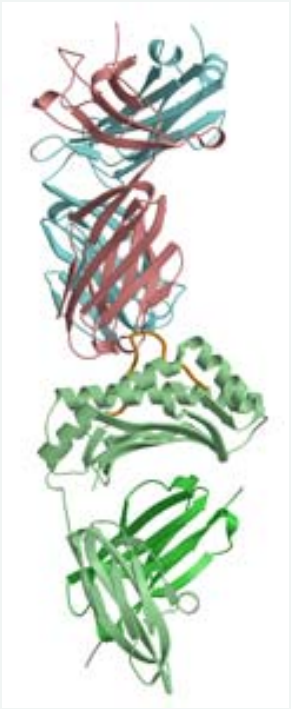

Dataset Packager
Dataset Image Directory Path
Browse...
Project Depositor
Repository Address: http://130.194.155.76:8080/fedora
Username: fedoraAdmin
Password: **********
Project Description File
/Users/steve/Desktop/DatasetTools/Example/2NW0.xml
Browse...
Project Upload Path
/Users/steve/Desktop/DatasetTools/Example/Upload
Browse...
Upload
Wrote to /Users/steve/Desktop/DatasetTools/Example/Upload/oai_ds.xml
Merged diffraction image metadata into single file for harvest.
5 file(s) to be uploaded...
Uploading.. 2NW0 Diffraction Images.tar.bz.0
Uploading.. 2NW0 Diffraction Images.tar.bz.1
Uploading.. 2NW0 Diffraction Images.tar.bz.2
Uploading.. oai_ds.xml
Uploading.. RELS-EXT.xml
Finished uploading files to Fedora Object changeme:10
Project URL:
http://130.194.155.76:8080/fedora/get/changeme:10
/Users/steve/Desktop/2NW0 Upload
Before upload.

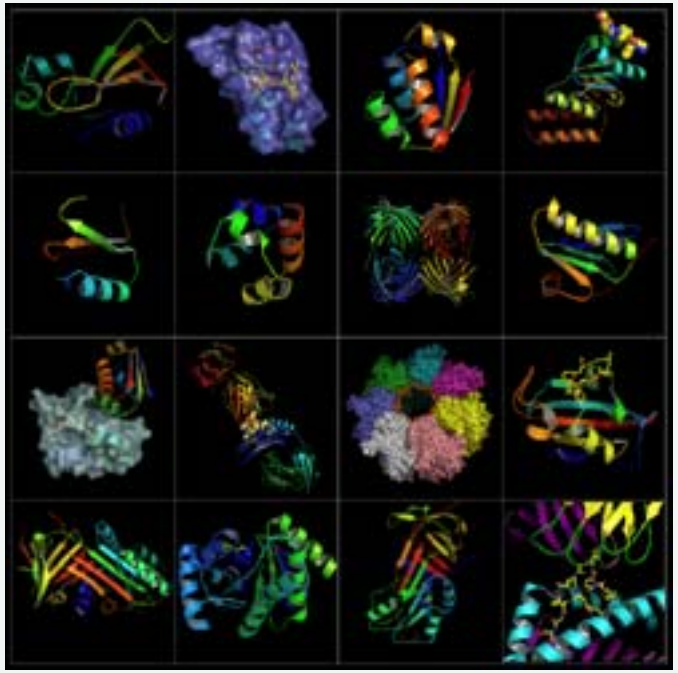
ARCHER Deployment: National Breast Cancer Foundation



ARCHER Deployment: ARCS Data Fabric

Coming ARCHER Deployment: Protein Crystallography





What researchers can expect from ARCHER

- A place to collect, store and manage experimental data
- Software tools focused on management of data & information
- Standardised and secure method of storing, accessing, and analysing research results
- Easier collaboration and sharing of research datasets & information

Future of ARCHER

- Currently testing the tools for release by late September
- Expecting that the partners will continue to develop the tools they created
- New enhanced versions already being worked on
- Looking at how these tools might be used within ANDS (Australian National Data Service) & ARCS (Australian Research Collaborative Service)

For more information…

Contact:

Anthony Beitz
ARCHER Portal & Dataset Product Manager
Ph: +613 9902-0584
Anthony.Beitz@its.monash.edu.au

See:

ARCHER Website: http://archer.edu.au
Demos: http://www.archer.edu.au/demo/