

DART

developing toolkits for e-Research

Dr Jeff McDonell, DART Project Director

Presentation to QUESTnet 2006, Coolum in July

What is DART?

Dataset

Acquisition / accessibility / annotation

e-Research

Technology

Acknowledgements

DART is a **proof-of-concept** project funded by DEST - *Department of Education, Science and Training*, to support collaborative research in Australia through MERRI - *Managed Environment for Research Repository Infrastructure*

What is DART trying to achieve?

- To develop tools to handle all the data and information management requirements of the **complete research lifecycle**
- To collect and manage **large datasets**, associated with sensor networks, synchrotrons, telescopes, etc.
- To support **collaborative research** and **annotation** needs
- To deal with **intellectual property**, **privacy** and **security** issues
- To create **customised portals** for research demonstrators
- To handle research **publication**, **discovery** and **access**

or to put it another way.....



What should researchers expect from DART?

- A place to store data – not in the lab or home PC or DVD disk!
- Useful tools for them to use in their everyday research
- Software tools focused on the management of data and information
- Potentially a useful portal applicable to their research field
- A standardised and secure method of storing, accessing, analysing and annotating their research results
- Easier to collaborate, share information and then publish results

DART Design Criteria

- Identify the best-of-breed solutions in each area
- Use Open Standards
- Choose open-source software wherever possible
- Leverage existing work and expertise - don't **reinvent the wheel**
- Identify common frameworks for:
 - security
 - network transport
 - Integration across all DART work packages

DART Research Demonstrators

The use of demonstrators is designed to show the value of an end-to-end DART lifecycle approach, for both scientific and non-scientific research

The three research areas chosen as DART demonstrators are:

- X-Ray Crystallography
- Climate Research
- Digital History

Monash, Queensland and James Cook University researchers are involved in all 3 demonstrators testing the 'proof-of-concept' DART objectives.

How do we build demonstrators?

The DART demonstrator tasks are to:

- Engage with suitable researchers at each partner university for each of the three chosen research disciplines
- Define the research activities applicable to each research area
- Embed Information Management specialists into the research teams
- Construct a custom designed prototype DART portal, incorporating research discipline specific software applications
- Progressively refine the model as the DART project adds new features and support

DART logistics

DEST funding of A\$3.235 million involving:

- 3 partners: - Monash University (host) in Melbourne
- University of Queensland in Brisbane
- James Cook University in Townsville
- 5 areas of focus within the DART work packages (WPs)
- 7 Chief Investigators
- 18 month project (expected to be completed mid 2007)
- 27 Separate DART work packages
- 40+ project team members!!

DART Chief Investigators



Andrew Treloar
(Project Architect)



Asad Khan



David Abramson



Ann Monotti



Jane Hunter

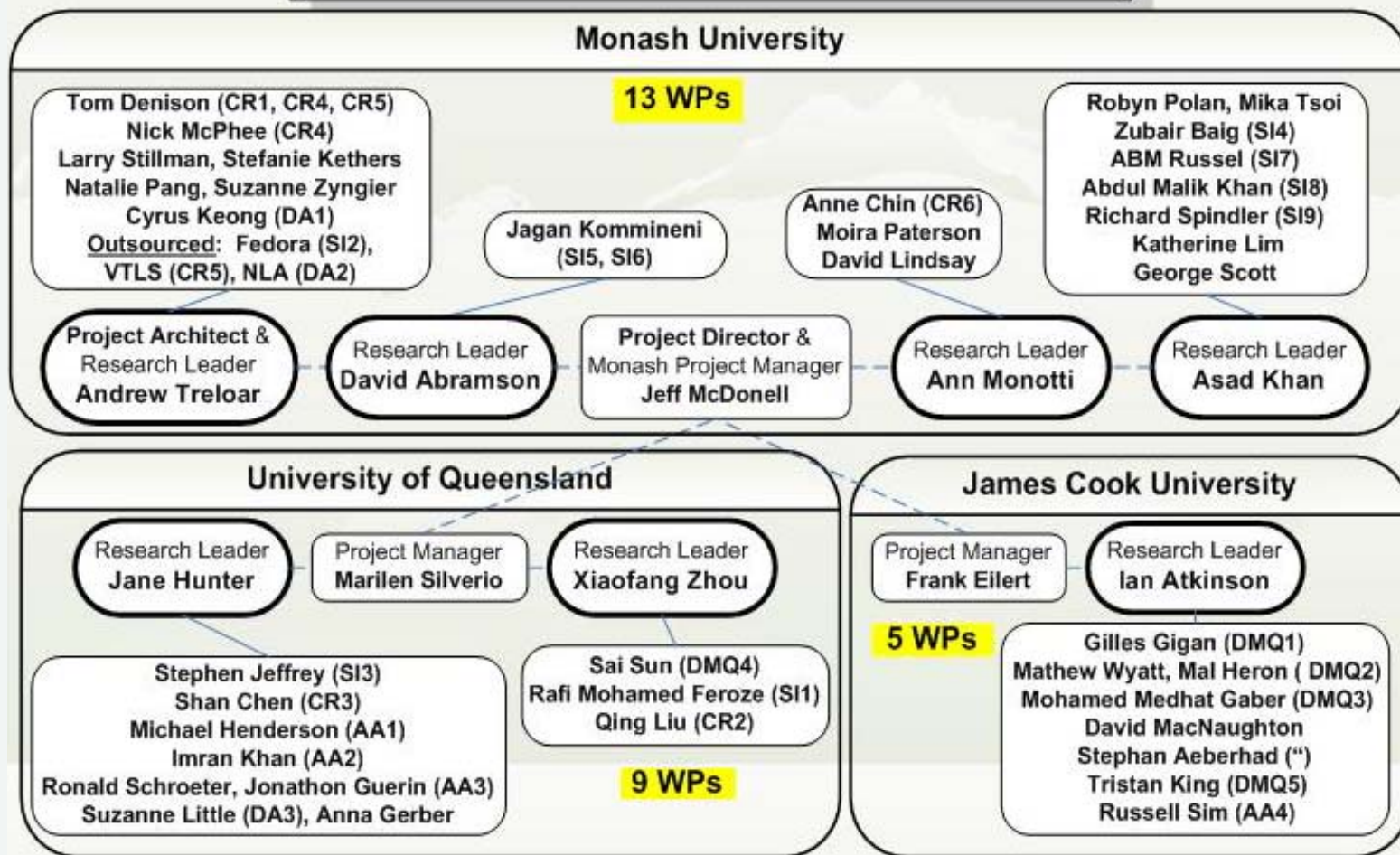


Xiaofang Zhou



Ian Atkinson

DART Project Organisational Structure



DART Work Packages

The 27 work packages that comprise DART cover five areas:

- Data Collection, Monitoring and Quality Assurance
- Storage and Interoperability
- Content and Rights
- Annotation and Assessment
- Discovery and Access

Data Collection, Monitoring and Quality Assurance

Developing front end research processes

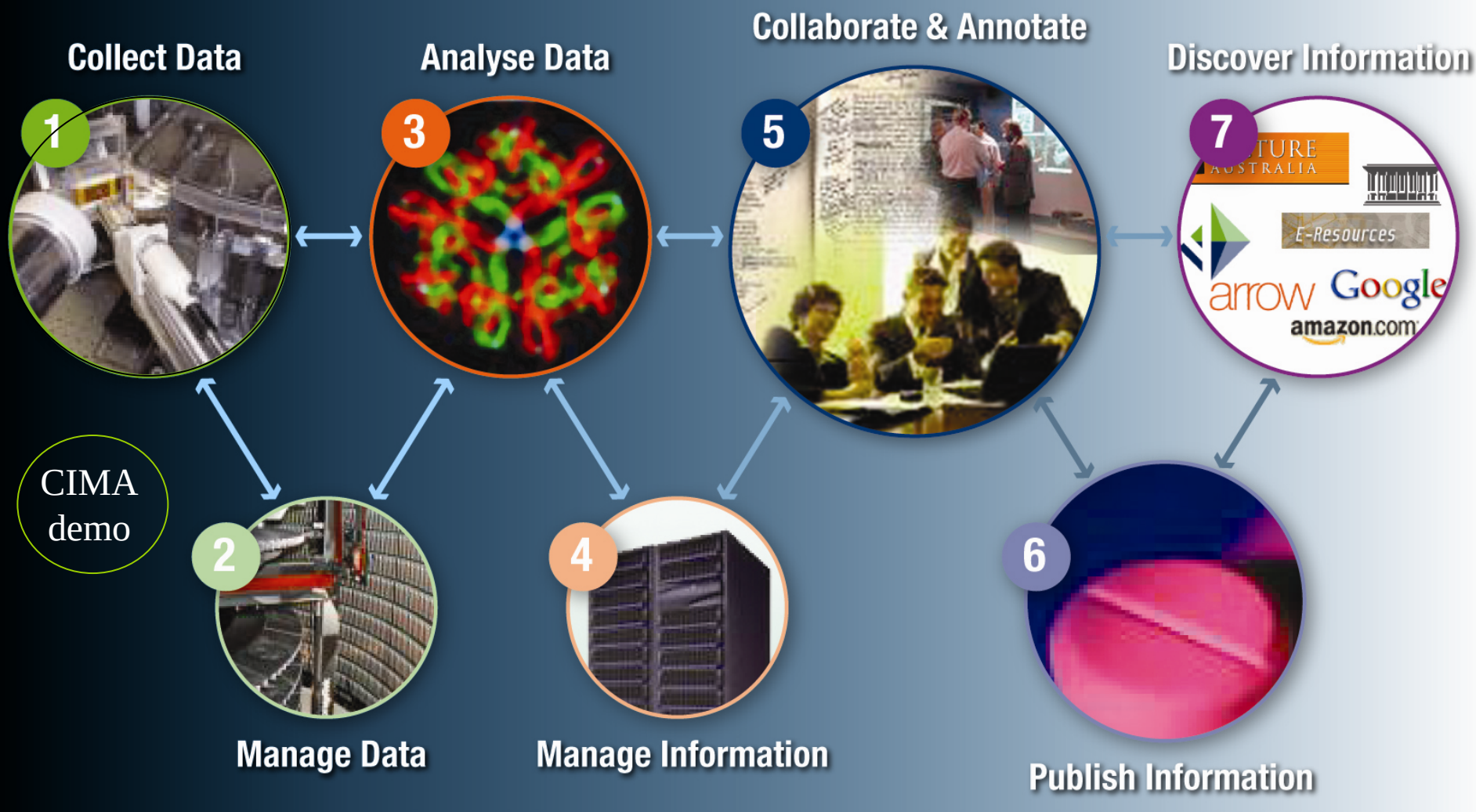
- 1: Connect instruments and sensors effectively to the network
- 2: Connect instruments to repositories using Storage Resource Broker (SRB) via Common Instrument Middleware Architecture (CIMA)
- 3: Ensure data is of sufficient quality to warrant curation
- 4: Online, remote access to working instruments & sensors
- 5: Improve intelligence of the storage framework

Ian Atkinson (JCU)

1, 2, 3, 5

Jane Hunter (UQ)

4



Storage and Interoperability

Developing middleware tools

- 1: Facilitate distributed data management with Fedora
- 2: Improve interoperability between SRB and Fedora
- 3: Support richer metadata to enhance discovery
- 5: Support a range of data replication systems - SRB, Globus & GFarm
- 6: Allow simulation data to be retrieved or dynamically regenerated

<i>Xiaofang Zhou (UQ)</i>	1
<i>Andrew Treloar (Monash)</i>	2
<i>Jane Hunter (UQ)</i>	3
<i>David Abramson (Monash)</i>	5, 6

Storage and Interoperability (con't)

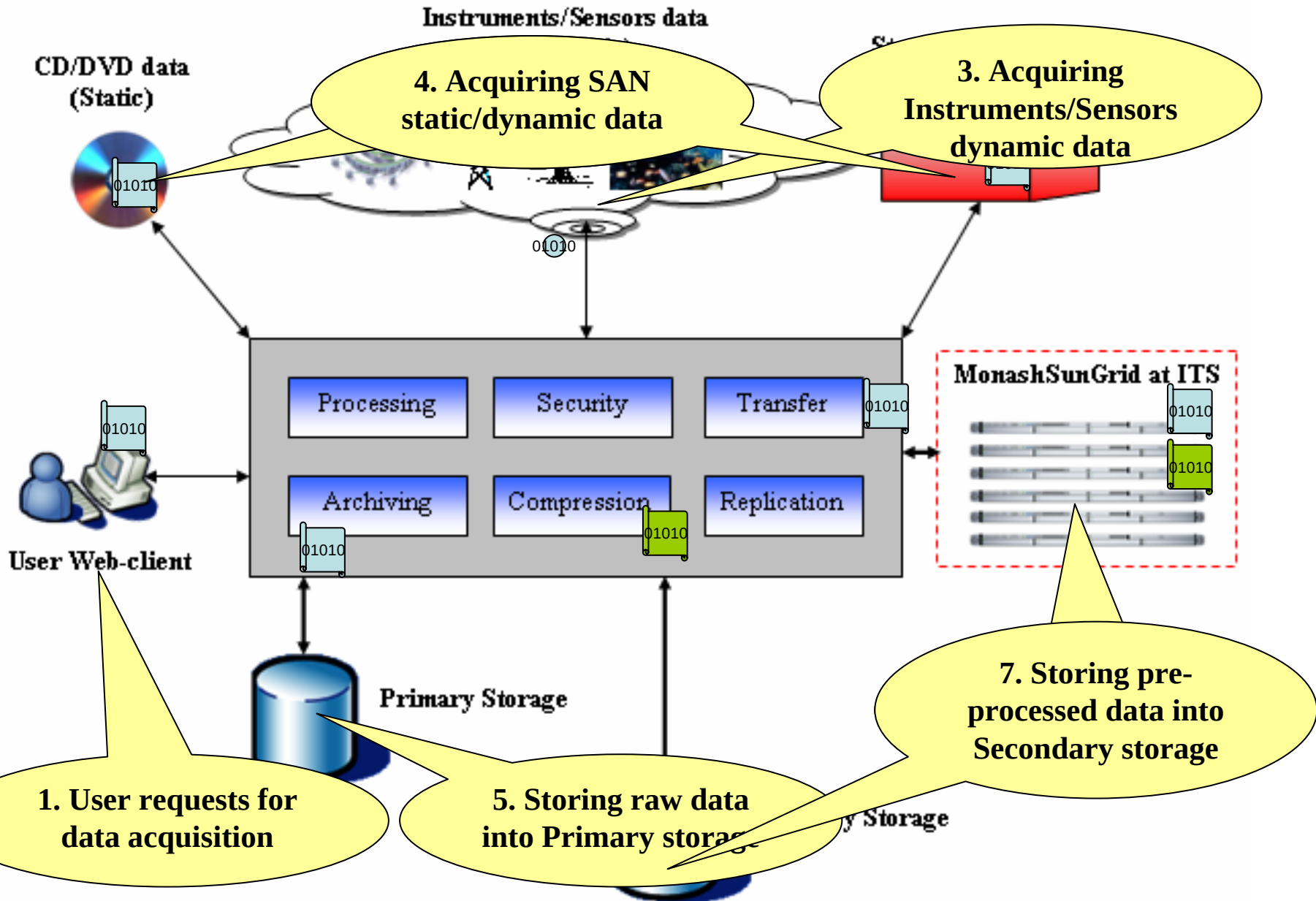
Developing secure data transfer and storage

- 4: Secure service for transferring data from instruments & sensors to repositories via the Grid
- 7: Develop cost-effective data pre-processing system for secondary storage
- 8: Pilot long-distance high speed and secure data transfers between repositories
- 9: Scope and pilot storage infrastructure requirements

Asad Khan (Monash)

4, 7, 8, 9

2. Manage Data



User Web Client interface for data pre-processing



Welcome to the Monash Sun Grid Central Portal at ITS

GridSphere

Home

Welcome to the Monash Sun Grid at ITS

This is the Monash Sun Grid central portal at Information Technology Services (ITS) division of Monash University. Before submitting a secured Grid job on Monash Sun Grid at ITS, you might want to view the status of the Grid first.



Monitoring Monash Sun Grid Resources at ITS

Grid
status
check

User login into the
web interface

Login

User Name
Password

☐ Remember my login

Login

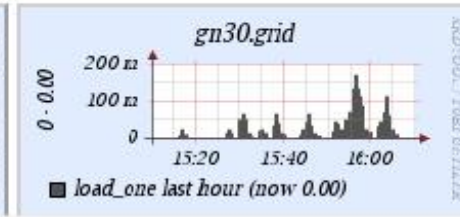
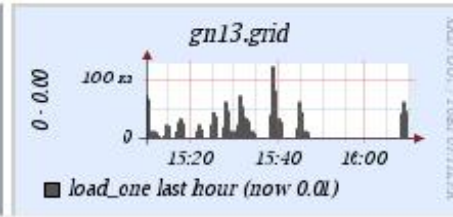
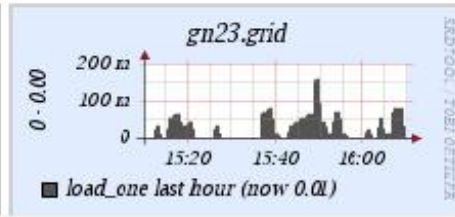
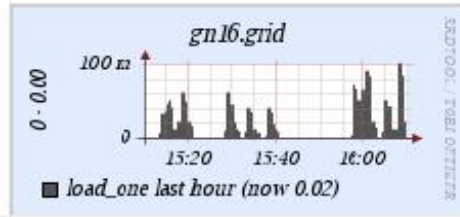
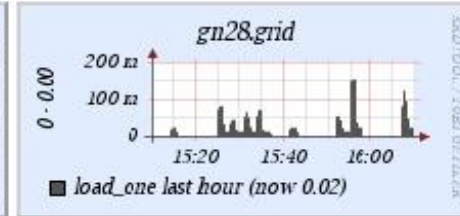
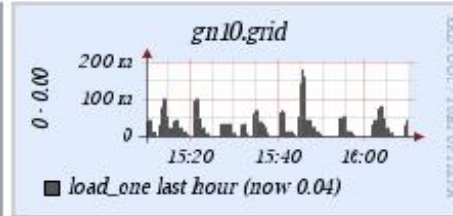
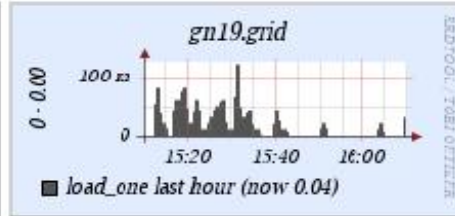
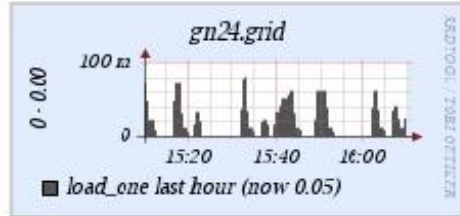
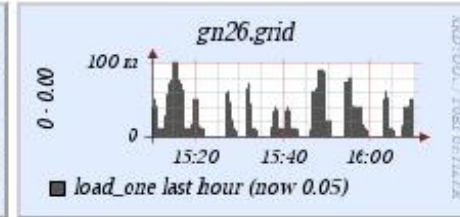
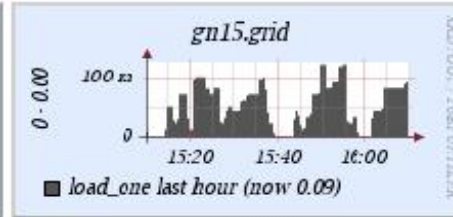
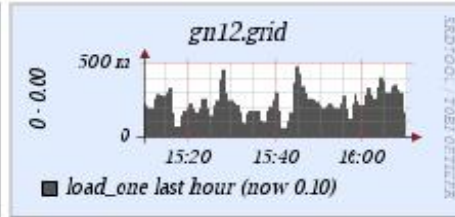
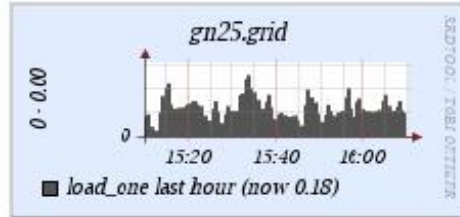
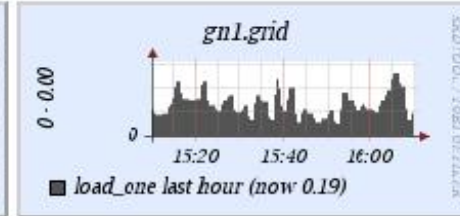
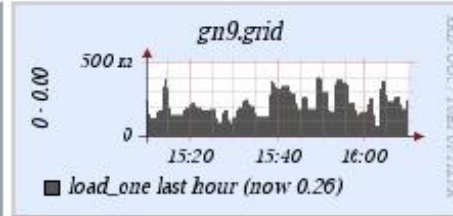
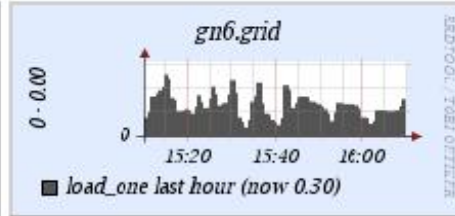
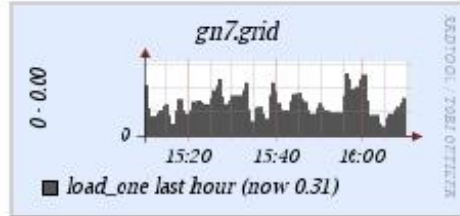
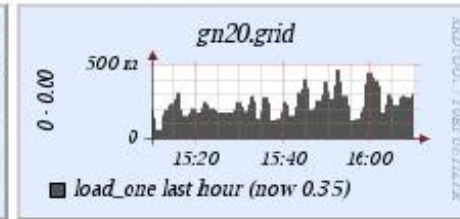
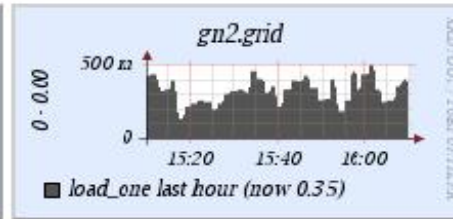
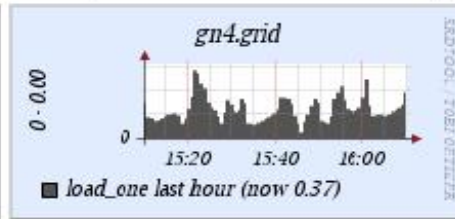
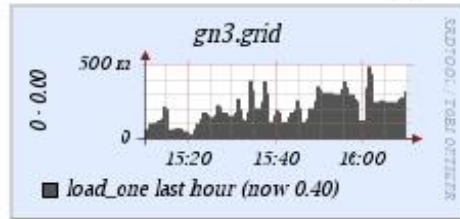
[Forget your password?](#)

powered by gridsphere

Last updated by A.B.M. Russel, Software Specialist (E-Research) on Monday 20 March 2006
Copyright 2006 Monash Sun Grid at Information Technology Services (ITS) Division, Monash University. All rights reserved.

Grid status display

Show Hosts: yes ☒ no ☐ | MonashSunGrid **load_one last hour sorted descending** | Columns 4





Welcome to the Monash Sun Grid Central Portal at ITS

Welcome Administration **Grid**

Registry Credentials Resources Files Jobs

? Job Submission Portlet

<<Previous

Next>>

Cancel

Step 1. Specify the application you would like to execute.

Application Details

Description MyJob

Executable file:///bin/date

Browse

single

Type of executable

Directory

Browse

Stdout file:///nfs/0/home/russel

Browse

Stderr file:///nfs/0/home/russel

Browse

Arguments

User sets job submission parameters



Welcome to the Monash Sun Grid Central Portal at ITS

Welcome Administration **Grid**

Registry Credentials Resources Files **Jobs**

Job Submission Portlet

<<List Jobs

Refresh View

New Job

Copy Job

Stop Job

Delete Job

Generic Application

Job Id	https://hn1 .its.monash.edu.au: 47207/26314/1143003126 /	Job Resource	hn1 .its.monash.edu.au	Date Submitted	Wednesday, 22 March 2006 03:52:07 PM EST
Job Description	MyJob	Job Scheduler		Last Changed	Wednesday, 22 March 2006 03:52:07 PM EST
Job Status	Job submitted	Job Queue		Date	

Job String

Job submitted to the
Grid for pre-processing



Welcome to the Monash Sun Grid Central Portal at ITS

Welcome Administration **Grid**

Registry Credentials Resources Files Jobs

? Job Submission Portlet

<<List Jobs

Refresh View

New Job

Copy Job

Delete Job

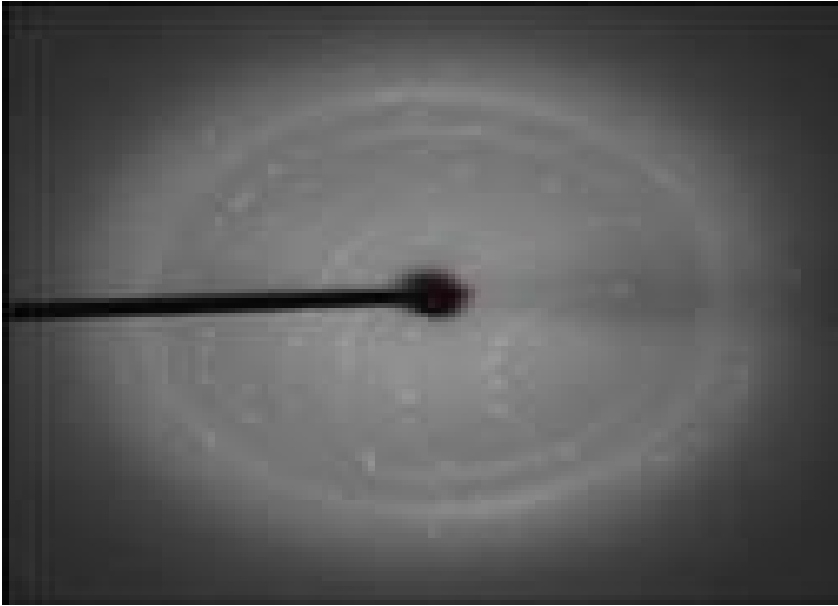
Generic Application

Job Id	https://hn1 .its.monash.edu.au: 47207/26314/1143003126 /	Job Resource	hn1 .its.monash.edu.au	Date Submitted	Wednesday, 22 March 2006 03:52:07 PM EST
Job Description	MyJob	Job Scheduler		Last Changed	Wednesday, 22 March 2006 03:52:07 PM EST
Job Status	Job completed with message success	Job Queue			Wednesday, 22 March

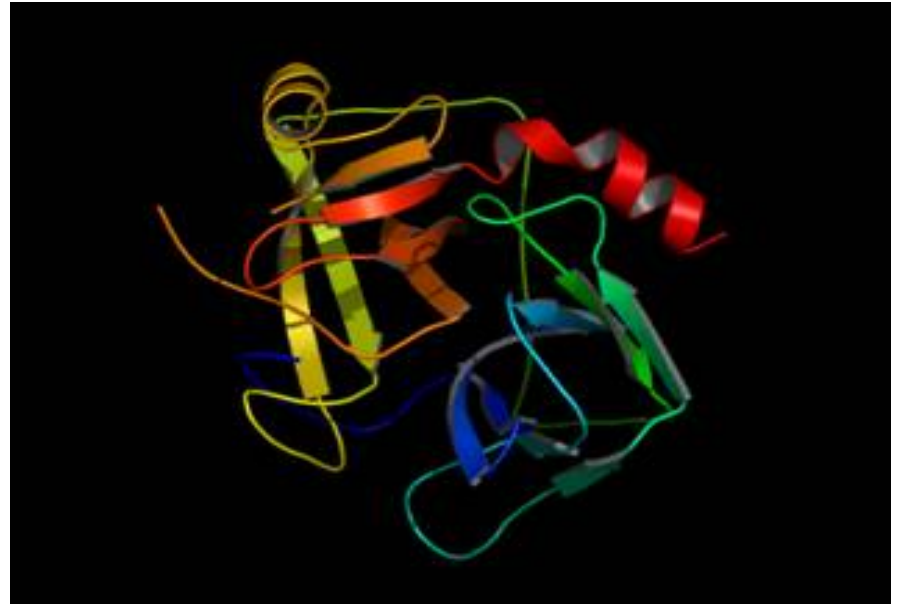
Job String

Job completed, output
saved on Secondary
storage

Sample Data Pre-processing Job Output



Protein crystallography raw data



**3D atomic structure of protein
after processing**

Content and Rights

Collecting data sources into institutional repositories (IRs)

- 1: Move data from personal repositories to secure trusted alternatives
- 2: Reduce barriers to content acquisition by rights assignment for non-Science researchers
- 3: As above – but for Science researchers
- 4: Improve management practices in research communities
- 5: Assist researchers deposit datasets and other digital objects into IRs
- 6: Clarify legal issues around IP, information security and privacy

Andrew Treloar (Monash)

1, 4, 5

Xiaofang Zhou (UQ)

2

Jane Hunter (UQ)

3

Ann Monotti (Monash)

6

Annotation and Assessment

Including collaboration tools for research

- 1: Allow researchers to annotate each other's work
- 2: Improve annotation and deposit rates by allowing end-user control
- 3: Help annotation services contribute to the life and productivity of research communities
- 4: Foster emerging wiki-based collaborative work practices in research teams

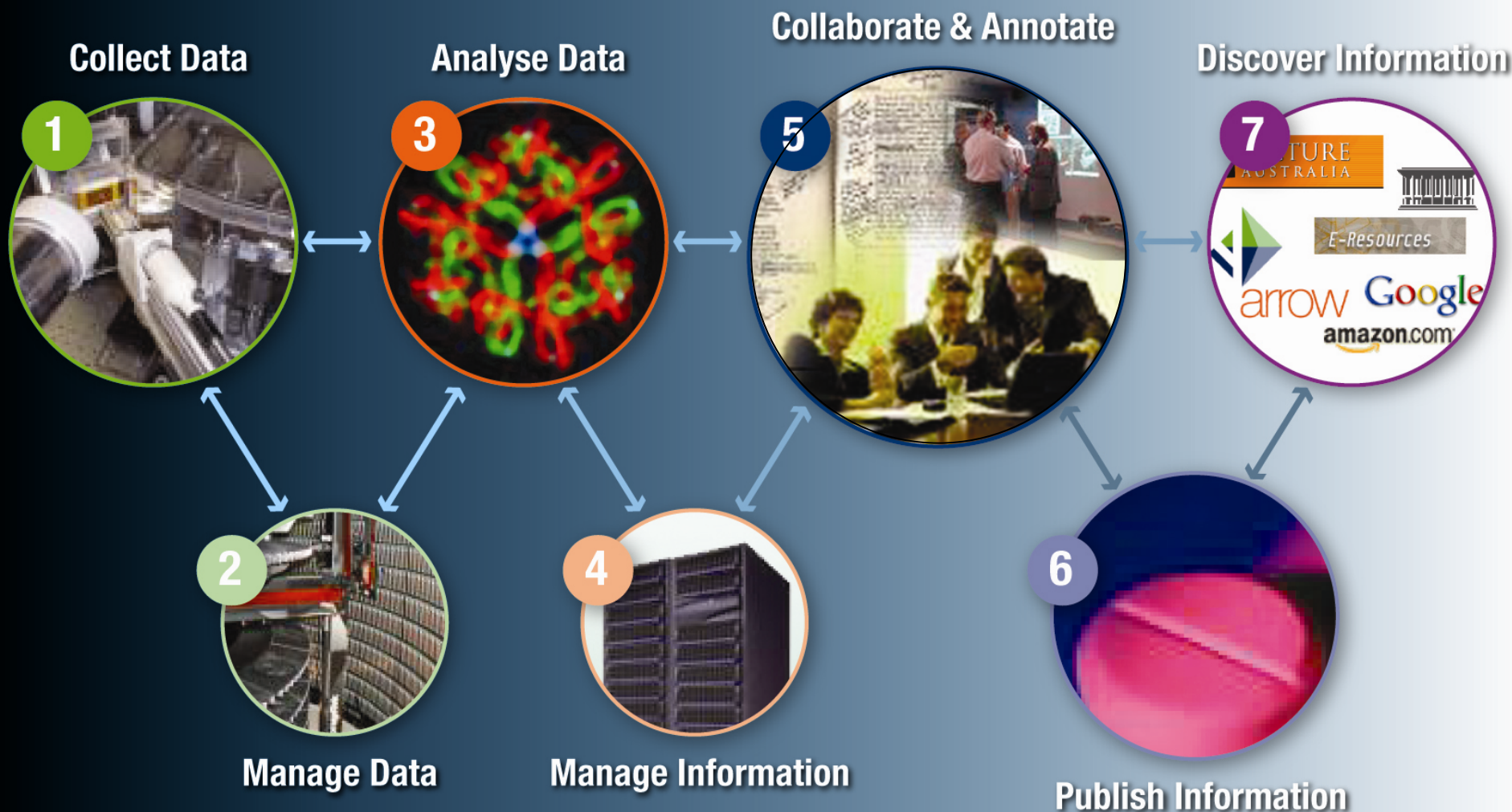
Jane Hunter (UQ)

1, 2, 3

Ian Atkinson (JCU)

4

← data ← information →



Discovery and Access

Searching, browsing and discovering resources

- 1: Improve repository deposit rates, sharing and reuse by allowing end-user access
- 2: Improve repository deposit rates, sharing and reuse by improving discoverability
- 3: Reduce effort for creating metadata schemas and improve interoperability

Andrew Treloar (Monash)

1, 2

Jane Hunter (UQ)

3

Discovery examples



PICTURE AUSTRALIA

Whose images? About us Contact us FAQs Site map News Ordering images Links

NEWS

Monash University's Centre for Gippsland Studies joins PictureAustralia - [\[more\]](#)

Swinburne University of Technology images temporarily unavailable - [\[more\]](#)

Looking for images of Australian? PictureAustralia® is the place to go. Search for people, places and events in the collections of libraries, museums, other cultural agencies, in Australia and abroad - all at the same time and order quality prints at your leisure.

Search for images:

[Advanced Search](#) [Browse](#) | [Search Help](#)

Click and Flick - PictureAustralia needs your pics!

Email your favourite images...



Indigenous Australians are advised that PictureAustralia may include images or names of people now deceased.

Interested in making your image collection available through PictureAustralia? Check out the [For Contributors](#) pages to see what's involved.

Last update: 12 May 2006 | pictaust@nla.gov.au | © Registered Trade Mark



List of titles

[Simple search](#) [Advanced search](#) [Expert search](#)

Search query

Select a database...

VITAL Repository

Showing 1 - 15 of 83 entries.

[« Previous](#) | [1](#) [2](#) [3](#) [4](#) [5](#) [6](#) | [Next »](#)

Understanding Safety Performance Using Safety Climate And Psychological Climate

Creator	Clissold, Gemma
Description	This paper aims to explore the relationship between psychological climate, safety climate and safety performance. Safety research has moved from a central focus on safety specific explanations of safety performance, to encompass more general management principles (performance management). This research aims to contribute to this body by exploring the way in which psychological climate affects safety performance. This paper compares the fit
Subject	Occupational health and safety
Subject	safety management
Subject	safety climate
Subject	safety performance
Subject	psychological climate

The Organizational Culture Profile: an Australian Perspective

Creator	Sarros, James C.
---------	------------------



[Whitlam Dismissal](#)

DART Demonstrators

X-Ray Crystallography

- Focussing on diffractometers and protein crystallography
- Using CIMA instrument interfaces
- James Whisstock (MU), Jenny Martin (UQ) and Ian Atkinson (JCU) major researchers involved

Climate Research

- Focussing on ocean and atmospheric data – example of merging Great Barrier Reef data around Heron Island to predict weather
- Amanda Lynch (MU) and Stuart Kininmonth (AIMS) early adopters

DART 3rd Demonstrator

Digital History

- 3 key projects with a Humanities focus
- 'Gugu Badhun Digital History' project at JCU
- 'Women on Farms' project at MU
- 'Western Cape Community Agreement' project at UQ
- Dealing with video storage and annotation, survey data, authorisation and security, community involvement, etc.

DART Deliverables



DART Deliverables

By mid 2007, DART aims to provide:

- Working **proof-of-concept** demonstrations
- Feedback from researchers in the three demonstrators
- Customised portals for specific demonstrator research disciplines
- Assessment of the value of the DART integrated lifecycle approach
- A clear understanding of how to turn **proof-of-concept** into robust **production-ready** systems
- Reports recommending best practice in several areas

Note that the DART project **WILL NOT** deliver production services (yet)!

How has DART progressed?

Fast startup:

- Started Dec 2005, DART now has 40+ staff and researchers on board

Collaborative Project:

- 27 WPs in 3 partner universities: Monash, Queensland, James Cook

Effectively managed:

- 7 Chief Investigators, strong Project Office and Board of Management

Grounded in research practice:

- Building 3 demonstrators with research teams from the 3 partners

Common standards used to develop generic IT tools:

- Fedora, GridSphere, SRB, XACML, Shibboleth, Annotea, CIMA, etc.

DART Key Achievements

1. Strong progress in data capture and instrument integration
2. Investigating storage and replication of very large datasets (up to Petabytes) across diverse networks
3. Embedded Information Management specialists into key research teams, to address data and information management requirements
4. Developing annotations software for 3-D models, video and audio
5. IP and privacy are being reviewed by a Law Faculty
6. Investigating Creative Commons and Science Commons licensing
7. Working to utilise Shibboleth and Grid security standards
8. Developing search tools, metadata schema registry, wiki tools, etc.

DARTs in use by the ARCHER?

ARCHER is a new DEST project that is:

- Funded for \$4.545 million in 2007
- To take the **proof-of-concept** outcomes of DART and turn them into **production-ready** ARCHER software tools, that are...
- Customised through dedicated task forces to suit the needs of the 9 NCRIS priority research capabilities, plus
- Two specialised task forces for the Humanities and Social Sciences

Useful DART tools for ARCHER (1)

Compute/storage:

- Interface to instruments/sensors, including CIMA video record ('lab book')
- Interface to distributed computing (HPC / Grid)
- Interface between the hardware and the DART software
- Secure access to large scale data storage and repositories (SRB/Fedora)

Data Quality:

- Pre-analysis of data to automatically detect faulty/degrading data
- Seamless replication of data for backup and disaster recovery
- Support for multiple data replication systems (SRB / GFarm / Globus)
- Transfer of large datasets between systems efficiently and fault tolerantly

Useful DART tools for ARCHER (2)

Software tools:

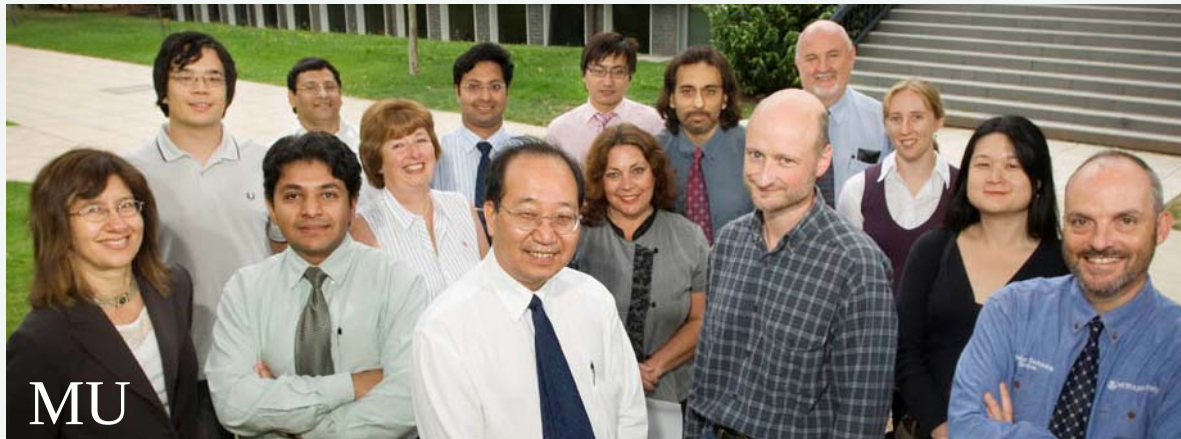
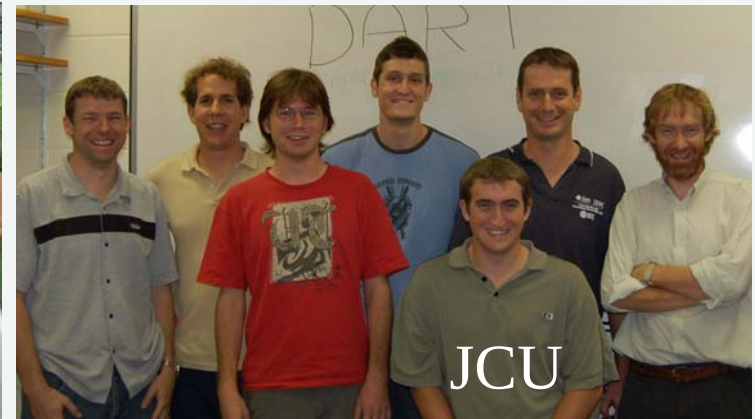
- Manage metadata, including defining, storing, searching, etc.
- Time saving workflow tools (Kepler)
- Manage authentication/authorisation and data security
- Deal with Science/Creative Commons licensing, IP and privacy issues
- Provide secure annotations for documents, datasets, video, audio, etc.
- Collaborate using Research-centric wiki & weblog communication tools

Usability :

- GridSphere portal to tie everything together
- Migrate data from personal to institutional storage
- Support for legacy applications
- Interface to discovery/search engines

Acknowledgements

Without the hard work of all these people, DART would just not happen!



DART Contacts

- Web: dart.edu.au
- Email: dart@dart.edu.au
- Project Director: Jeff.McDonell@its.monash.edu.au
 - Phone +613 9905 4187
- Project Architect: Andrew.Treloar@its.monash.edu.au
 - Phone +613 9905 1138

Questions?