ARChER
Building data and information management tools for the complete research life-cycle

July 2006
What is ARCHER?

ARCHER is a new 2007 project now funded by DEST under SII funding to collaboratively build research-specific tools to handle all research data and information management requirements, building on the success of the DART and ARROW projects.
What can researchers expect from ARCHER

- A place to collect, store and manage data digitally
- Handy IT tools to use in their every day research
- Software tools focused just on management of data & information
- Adaptable portal applications relevant to their research field
- Standardised and secure method of storing, accessing, analysing and annotating research results
- Easier to collaborate, share information and publish results
What is ARROW?

Australian Research Repositories Online to the World

ARROW is a successful project funded by a DEST Systemic Infrastructure Initiative (SII) grant to identify and test software and solutions to support best practice institutional digital repositories, for example covering:

- e-Prints
- Digital Theses
- Electronic Publishing, etc.
What has ARROW achieved?

Since early 2004, ARROW has developed institutional repositories for digital research outputs. ARROW has:

- Designed and built commercial VITAL software with VTLS (US)
- Funded Open Source ARROW Web Services for all to use
- Populated repositories and developed the ARROW Discovery Service with the National Library of Australia (NLA)
- ARROW is now in use at eight Australian universities and the NLA

ARROW II is a 2007 project also now funded by DEST to support the Research Quality Framework (RQF) through ARROW
What is DART?

**Dataset**

**Acquisition / Accessibility / Annotation**

**e-Research**

**Technology**

DART is a successful *proof-of-concept* project funded by DEST to build tools to handle all the data and information management requirements for the complete end-to-end research life-cycle.
How has DART progressed?

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

Collaborative Project:
- 27 work packages 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 multiple research teams from the 3 partners:
  - X-Ray Crystallography
  - Climate Research
  - Digital History

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
How will the ARCHER project progress?

- ARCHER will create task forces to work collaboratively with existing research development teams, to build research-specific tools to handle all data & information management requirements.
- ARCHER will take ‘proof-of-concept’ tools developed by DART and ARROW and turn them into ‘production-ready’ software.
- ARCHER will apply them to specific research capabilities, tailor-making portal-based solutions adaptable to each research capability.
- But ARCHER is **NOT** a project to provide backend infrastructure, such as grid computing, network access, SANs, etc.
Turning proof-of-concept into production-ready

- Need to engage with the right researchers at universities & institutions
- Appoint research specific team leaders
- Add Information Management specialists into each team
- Design and build custom designed portals, as required
  - incorporating research specific software applications
- Progressively refine the DART and ARROW tools
- Extensive quality testing and usability trials
ARCHER model

- Research specific ARCHER Task Force
- Portal Development and Design
- Security and Legal Issues
  - Collaborative development of data and information management tools for NCRIS priority research capabilities
  - Software Tools
  - Instrument Interfaces (as required)
- Hardware Integration
  - Hardware Infrastructure (grid, cpu, storage, network)
Turning data into information

Protein crystallography raw data 3D atomic structure of protein after processing
Useful DART tools for ARCHER (1)

Compute/storage:
- Interface to instruments/sensors, CIMA video and Kepler workflow
- 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.
- 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.

Usability:
- GridSphere portal to tie everything together
- Migrate data from personal to institutional storage
- Support for legacy applications
- Collaborate using Research-centric wiki & weblog communication tools
Proposed ARCHER Deliverables

By the end of 2007, deliver production-ready software to handle data and information management requirements for specific research areas, tailor-made to their needs.

Note: To provide long term sustainable production services, ongoing support and service to the ARCHER production-ready software must also be provided - possibly through organisations like AARNet, CSIRO, APAC or similar.
Summary

- The DART project team has the expertise to implement ARCHER
- ARCHER will be industrial-strength DART for all researchers to use
- DART and ARROW have demonstrated experience in building software tools – some are already in production
- ARCHER will not be the owner of these tools, but they will be passed onto long-term service providers for sustainability & support
DART Acknowledgements
These are the DART people making things happen!