Taking aim and mostly hitting our targets: from DART to ARCHER (and beyANDS!)

Dr Andrew Treloar
DART Project Architect
ARCHER Director and Chief Architect
ARROW Technical Architect

Presentation to Clever Collections 2007
First some context

Grid

E-Scientists

Entire E-Science Cycle
Encompassing experimentation, analysis, publication, research, learning

Digital Library

Local Web

Publisher Holdings

Institutional Archive

Peer-Reviewed Journal & Conference Papers

Preprints & Metadata

Technical Reports

Virtual Learning Environment

E-Scientists

Graduate Students

Undergraduate Students

E-Experimentation

E-Scientists

Data, Metadata & Ontologies

Certified Experimental Results & Analyses

Preprints & Metadata

E-Scientists

Technical Reports

Peer-Reviewed Journal & Conference Papers

Source: Adapted from Liz Lyons, eBank UK Presentation
DART

- DART was a proof-of-concept project funded by the Department of Education, Science and Training (DEST) to support collaborative research in Australia.
- The funding was provided through the Systemic Infrastructure Initiative as part of the Commonwealth Government's Backing Australia's Ability - An Innovation Action Plan for the Future.
- DART stands for:
  - Dataset
  - Acquisition/Accessibility/Annotation
  - e-Research
  - Technologies
DART logistics

- 3 partners:
  - Monash University (lead) in Melbourne
  - University of Queensland in Brisbane
  - James Cook University in Townsville
- 5 technical areas of focus within the DART work packages (WPs)
- 18 month project, finishing in June 2007 (mostly)
- 28 Separate DART work packages
- 40+ project team members
- Led by Dr Jeff McDonell
What did DART try to achieve?

- To develop software tools to handle the data and information management requirements of the complete research lifecycle
- To collect and manage large datasets, associated with instruments, such as sensor networks, X-ray diffractometers, 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......
DART VC Diagram

building the new collaborative e-research infrastructure

Collect Datasets

Collaborative Research Analysis

Annotation - Dynamic Documentation

'Publish'

'Discover'

Raw Data Store

Analysed Data Store

Data Collection Tools

Middleware Tools

Annotation Tools

Access Tools

Large Dataset Management Tools

Data Curation Tools

Large Dataset Management Tools

Data Curation Tools

Monash University

JCU

University of Queensland

Australia
DART work packages

- The work packages covered five broad technical areas:
  - Data Collection and Monitoring
  - Storage and Interoperability
  - Content and Rights
  - Annotation and Assessment
  - Discovery and Access
- Details on the website, including all project outputs:
  - reports, source code, documentation
DART Achievements

- Strong progress in data capture and instrument integration
- Investigated storage and replication of very large datasets across diverse networks
- Placed Information Management staff into research teams, addressing their data and information management requirements
- Developed annotation software for 3-D models, video and audio
- Reviewed IP and privacy around datasets and e-research practices
- Investigated Creative Commons & Science Commons licensing
- Worked to utilise Shibboleth, PKI and Grid security standards
- Developed search tools, metadata schema registry, Plone tools, etc
- Production deployment of X-Ray Crystallography tools
DART Demo

- NOTE: This is only a partial demo of the range of work performed in DART - for more see the website


But is it Collectible?

- Clever Collections need high quality data in them
- Ensuring high quality data needs
  - inbuilt quality control ✓
  - near-frictionless data collection from instruments ✓
  - (semi) automated metadata extraction ✓
  - good data management ✓
From DART to ARCHER

- Australian ResearCH Enabling enviRonment
- ARCHER is a DEST funded project for 2007/8 that is taking the proof-of-concept outcomes of DART, turning them into production-ready ARCHER software tools, and packaging them, along with other open-source components, for deployment
- These tools are being developed as modular middleware components, generalising from the needs of a number of designated National Collaborative Research Infrastructure Scheme (NCRIS) priority research capabilities
- Focus is on data-centric collaboration
- ARCHER aims to have an internal prototype available in mid December 2007, final versions mid 2008
ARCHERTecture

**Web Frontend**

**ARCHER Components**

- Workflow Invocation (Kepler, Taverna)
- Guest Access Capability (AAF)
- Data Collection & Instrument Mgt (DIMSIM)
- Dataset Deposition
- Metadata Creation
- Offline Object Management (YourSRB)
- Collaborative Workspaces (Plone)
- Collaborative Annotation (TBD)
- Data Export
- Metadata Management

**Stores**

- Ontology Schema Registry
- Metadata Schema Registry
- Collaborative Workspaces
- Object Storage (SRB)

**Enabling Infrastructure**

- VO Management (MAMS)
- Authentication (AAF)
- Workflow (Kepler)
- Grid Computing
ARCHER contraction of scope

Grid

Entire E-Science Cycle
Encompassing experimentation, analysis, publication, research, learning

Source: Adapted from Liz Lyons, eBank UK Presentation
ARCHER into ANDS

- ARCHER is currently re-aligning itself with ICI and ANDS
- ARCHER will be applying for NeAT funding for two separate projects in FY2008/9
  - Collaborative Workspaces
  - Data Acquisition and Management
- Regardless of these projects, outcomes of ARCHER likely to be useful in ANDS Repositories program
ANDS and Collections

- ANDS will need to rely on well-curated data collections to meet the needs of the Australian Data Commons
- ANDS assumes that data is well managed in institutionally-supported/discipline-supported repositories
  - Like those being produced by ARROW, ARCHER and APSR
- ANDS will present data through a variety of discipline-focused collections
- One of the ANDS Utilities is likely to be a Collections Registry
- For more read *Towards the Australian Data Commons*
  - Or just come and hear Rhys tomorrow at 3.45!
Questions?

- Web:
  - dart.edu.au
  - archer.edu.au

- Email:
  - andrew.treloar@its.monash.edu.au

- Funding for DART/ARCHER has been provided through the Systemic Infrastructure Initiative as part of the Commonwealth Government's Backing Australia's Ability - An Innovation Action Plan for the Future.