Outcomes of the NeAT Program: eResearch tools and services for national research communities

Dr Paul Coddington
ARCS Projects Manager
Deputy Director, eResearch SA

Supported by:

NeAT Program

- Established under NCRIS Platforms for Collaboration
- Aimed at increasing the uptake of eResearch
  - Embed eResearch practice into research communities
  - particularly in NCRIS capabilities, HASS and other national research communities
  - Increase number of people with eResearch skills
- Develop new eResearch services and applications to meet requirements of particular research disciplines
  - but can potentially be reused in other areas or act as exemplars
  - preferably utilise or enhance the general-purpose PIC services
  - adopt, adapt, extend, harden existing tools or software
- Jointly funded and managed by ANDS and ARCS
- Similar to NeCTAR eResearch Tools program
NeAT

- National eResearch Architecture Taskforce (NeAT)
- Established by AeRIC
- Approval of (and feedback on) proposed projects
- Membership from experienced eResearchers

Rhys Francis (chair)  Jane Hunter
David Abramson      Marco La Rosa
Ian Atkinson        Steve Maddocks
Nathan Bindoff      John McGee
Markus Buchhorn     Deborah Mitchell
Adrian Burton       Ross Wilkinson
Paul Coddington     Tony Williams
James Daiziel       Robert Woodcock
Paul Davis          Lesley Wyborn
Ben Evans           John Zic
Lindsay Hood

NeAT Projects

- Funded equally by ARCS and ANDS
- Total of $12M over 4 years
- Scale of projects
  - Big enough to make an impact at a national scale
  - Small enough to enable projects across many discipline areas
  - Typically 2-4 EFT for 2-3 years for each project ($500K-$1M)
    - plus similar in-kind effort from eResearch service providers,
      research institutions, research communities
- Sustainable, long-term tools and services
- Projects had to specify organisations to provide ongoing
  operational support of tools and services
  - Institutions, eResearch providers, NCRIS capabilities
Governance

- DIISR and AeRIC
- ANDS and ARCS governance committees
- Project selection by NeAT
- Annual Reviews by NeAT Executive Committee
  - Members from AeRIC, ARCS, ANDS
- Each project was governed by a Project Committee
  - representatives from ARCS, ANDS, the research discipline (including end users), service operators, project manager
- Project manager for each project
- NeAT Projects Management Office (PMO)
  - ARCS Projects Manager
  - Projects Coordinator (Alan Glixman)
  - ARCS Business Manager, ARCS and ANDS admin staff

Selection Rounds

- Proto-NeAT
  - Several small-scale projects in 2007-08
  - Continuation of existing eResearch projects and/or staff
  - Astronomy, bioinformatics, characterisation, chemistry, grid tools, high-energy physics, video collaboration
- Round 1
  - Projects targeted at NCRIS capabilities and HASS
  - 6 projects selected (started July 2008 for 2-3 years)
- Round 2
  - Open call for proposals
  - 70 proposals received
  - 8 projects selected (started July 2009 for 2 years)
Round 1 NeAT Projects

- **ASeSS** - ASSDA Services for e-Social-Sciences (ASSDA)
- **Aus-e-Lit** (AustLit, Assoc. for Study of Australian Literature)
- **DataMINX** - Data in Microscopy Imaging Neutron X-ray (Characterisation)
- **DIAS-B** - Data Integration & Annotation Services in Biodiversity (ALA)
- **MACDDAP** - Marine and Climate Data Discovery and Access Project (IMOS)
- **SISS** - Spatial Information Services Stack (Auscope)

DataMINX subsequently split into multiple projects:
- **PIC-AMMRF** - Platforms for Collaboration in AMMRF (AMMRF)
- **ANDS projects** for metadata capture and ingest (ASF and ANSTO)
  - Capture of metadata into searchable metadata catalogs

**ASeSS**

- **ASSDA Services for e-Social-Sciences**
  - Australian Social Science Data Archive, Australian Data Archive
  - Authorized search, analysis and visualisation for social science data
  - Tools to increase automation of data ingest and access control
Aus-e-Lit

- Collaborative Integration and Annotation Services for Australian Literature Communities
  - Austlit, Association for the Study of Australian Literature
  - Full text search of documents held by Austlit
  - Federated search across multiple distributed databases as well as Austlit

- User annotations and creation of customized compound digital objects
- LORE tools - Literature Object Reuse and Exchange
DIAS-B

- Data Integration & Annotation Services in Biodiversity
- Atlas of Living Australia
- Metadata repository supporting registration, discovery and search
- Support for user annotations, e.g. for flagging problems with data quality

MACDDAP

- Marine and Climate Data Discovery and Access Project
- IMOS, Australian Oceans Data Network (AODN), TPAC Data Portal
- Add support for OGC spatial data standards and metadata standards
- Automation of data ingest and metadata generation in standard formats
SISS

- Spatial Information Services Stack
  - Auscope
  - Open source software stack for spatial data
  - Conforming to OGC standards
  - Focus on geoscience data sets but more generally applicable

DataMINX

- DataMINX - Data in Microscopy Imaging Neutron X-ray
  - NCRIS Characterisation – ASF, ANSTO, AMMRF
  - Moving data from instruments (synchrotron, ANSTO, microscopy) to data repositories and user’s desktop, authorized sharing of data
  - Automated capture of metadata into searchable metadata repository
PCA

- Platforms for Collaboration in AMMRF
  - Australian Microscopy and Microanalysis Research Facility (AMMRF)
  - Data management for AMMRF microscopy facilities
  - Web-based tool for mapping research requirements to facility capabilities

Round 2 NeAT Projects

- Auscover workflows (TERN Auscover)
- Aus-e-Stage (AusStage)
- BioFlows (BPA, Australian Bioinformatics Facility, QFAB, VLSCI)
- Biosecurity Collaboration Platform (AAHL and ABIN)
- Human Variome database (Human Variome Project)
- NCJRND - National Criminal Justice Research Data Network
- PODD - Phenomics Ontology-Driven Data (APPF and APN)
- Remote Computed Tomography (Australian Synchrotron and ANU)
Auscover Workflows

- Terrestrial Ecosystems Research Network (TERN) Auscover
- Workflow framework for processing satellite images
- Workflows for data providers and end users

Aus-e-Stage

- Collective Intelligence and Collaborative Visualisation for Creative eResearch
- AusStage performing arts database
- Analysis and mapping tools for AusStage performing arts database
- Tools for input from audiences (mobile phone app, SMS, web portal)
Aus-e-Stage

- Overlays of historical maps onto Google Earth

BioFlows

- Bioinformatics workflows
  - BioPlatforms Australia, Australian Bioinformatics Facility, QFAB, VLSCI
  - Easy-to-use web-based bioinformatics workflow tool (YABI)
  - Hosted locally on a customisable server appliance
  - Jobs run on local compute facilities or ARCS Grid
BCP

- Biosecurity Collaboration Platform
  - Australian Animal Health Laboratory (AAHL)
  - Visual telecollaboration services for biosecurity information

Human Variome

- Software and Data Support for the Australian Node of the Human Variome Project
  - Database for genetic variations for clinical studies of genetic diseases
NCJRDN

• National Criminal Justice Research Data Network
  • Enable access to criminal justice data sets held by state organisations

PODD

• Phenomics Ontology-Driven Data
  • Australian Plant Phenomics Facility and Australian Phenomics Network
  • Data management for plant and mouse phenomics using Fedora
Remote CT

- Remote Computed Tomography Reconstruction, Simulation and Visualisation
  - Australian Synchrotron and ANU micro-CT facility
  - Remote execution of computed tomography and 3D visualization

Summary of Outcomes

- Most projects (and proposals) were focussed on managing, accessing and sharing data
  - Other projects were on data processing and visualisation
- Developed or enhanced over 30 software tools
  - Open source, freely available
- Deployed eResearch services based on these tools
  - At multiple sites in many cases
  - Supported for at least a year after project completion
- Hundreds of data sets made (more easily) accessible
  - Standard interfaces, metadata, data formats, discovery
- Some tools already used in other research disciplines
  - e.g. three tools used in other NeAT projects
Summary of Outcomes

• Directly funded over 100 people from 20 institutions
  • Many more from >50 organisations contributed to the projects
• Thousands of researchers using the tools and services
  • Usage should grow for some more recently completed tools
• Research groups reporting that the tools are providing significant improvements to research practice
• Raised awareness of eResearch and PfC in research communities
  • 9 projects targeted needs of NCRIS capabilities
  • 4 projects in humanities, arts, social sciences
• Provide exemplars of eResearch practice
  • See the NeAT projects brochure

Learnings - Problems

• Collaboration is hard
  • and the more organisations in the project, the harder it is
• Finding and keeping people with expertise & experience
• External dependencies
  • Services, hardware, data, people, integration, hosting, …
• Getting promised co-investment
• Timely reporting
• Running the program jointly through two organisations
• Not using contracts meant more admin effort
• Many layers of governance meant sign-off was slow
• Culture – research vs development vs service provision
• Lost experienced people in the gap between NeAT and NeCTAR
Learnings

• Need significant resources for program management, administration, communications
• Really successful projects had
  • a good project manager with enough time commitment
  • good experienced senior IT staff
  • an engaged and knowledgeable steering committee
  • good engagement with end users
  • creative responses to problems
• Lessons learned provided as input to NeCTAR
• Useful to have different scales of projects
  • eResearch Tools Program
  • Virtual Laboratories Program
  • Research Cloud Node Program – application migration