

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

Paul Coddington

eResearch SA and ARCS, Adelaide, Australia, paul.coddington@ersa.edu.au

THE NATIONAL ERESEARCH ARCHITECTURE TASKFORCE

The National eResearch Architecture Taskforce (NeAT) was established by the Australian eResearch Infrastructure Council (AeRIC) under the NCRIS Platforms for Collaboration (PfC) capability. NeAT was responsible for identifying a portfolio of projects to develop and implement new eResearch methods, tools and services, which were jointly funded by the Australian National Data Service (ANDS) and the Australian Research Collaboration Services (ARCS) for a total of \$12M. NeAT was chaired by the Executive Director of AeRIC, included the Executive Directors of both ANDS and ARCS, and comprised a group of experts, appointed by AeRIC, to provide the necessary experience and expertise from a wide cross-section of relevant disciplines.

THE NEAT PROGRAM

NeAT projects were aimed at increasing the uptake and practice of eResearch, by developing eResearch tools and services that would meet the particular needs of a specific national research community or discipline and have significant impact on research practices in the target research community. Preference was given to tools that could potentially have broader use. Projects also needed to have short timeframes (1-2 years) for deployment of production-level tools and services; utilise or enhance the general-purpose PfC services where appropriate; adopt and adapt existing standards and tools where possible, particularly international developments, to create longer lasting tools and services; include activities to encourage uptake by researchers and embed the eResearch tools and services into research community practices; specify organisations that agreed to provide long-term hosting and support of the tools and services; and have substantive research community engagement and co-investment.

NeAT held two project selection rounds. The first round of projects began in 2008 from proposals developed in consultation with major national research communities, particularly NCRIS capabilities, fulfilling PfC's mandate to assist other NCRIS investments. The second round began in 2009, with projects chosen from responses to an open call for suggestions on activities that could lead to next-generation eResearch tools and services. The open call revealed significant demand for eResearch tools, with 70 proposals from a wide range of research disciplines. The ideas were pruned based on the selection criteria and refined with input from NeAT before project plans were developed.

The result was a set of projects adding eResearch capabilities to most of the NCRIS infrastructure investments, as well as a number of projects that engaged research communities that otherwise were not directly supported by NCRIS, including three projects working with national organisations in the humanities and social sciences.

NEAT PROJECTS

Fifteen projects targeting a broad range of research disciplines were selected by NeAT. Each project received between \$500,000 and \$1,000,000, funding 2.5 to 4 EFTs for 18 months to 3 years, with significant additional in-kind effort and resources provided by the project partners. The size of the funding to each project was a tradeoff between having enough projects to provide a breadth of impact across many research disciplines, and supporting each project to a level that enabled delivery of significant outcomes to a particular discipline.

Most of the projects (and the many other project proposals) were strongly focussed on managing, accessing or sharing data, with the others providing tools and services for analysis, processing and visualisation of data sets. The NeAT projects, the target research communities (in parentheses) and a very brief summary of the eResearch tools developed are listed below.

- **ASeSS** - ASSDA Services for e-Social-Sciences (Australian Social Sciences Data Archive)
Tools for authorized data ingest, curation, storage, search and analysis for ASSDA and its user community.
- **Auscover Workflows** - Workflow Services to Enable a Large-Scale Temporal-Spatial Ecosystem Digital Information Service (TERN Auscover)
Workflows for processing satellite imagery on compute clusters and grids, for both data providers and end users.
- **Aus-e-Lit** - Collaborative Integration and Annotation Services for Australian Literature Communities (AustLit; the Association for the Study of Australian Literature)
Federated search across multiple distributed databases; tools for creation, storage and search of customized compound digital objects; tools for creation, storage and search of annotations on digital objects.
- **Aus-e-Stage** - Collective Intelligence and Collaborative Visualisation for Creative eResearch (AusStage; Australasian Association for Theatre Drama and Performance Studies)

Tools for exploring collaboration networks and mapping information from the AusStage performing arts database; and gathering feedback from performances using mobile devices.

- **BioFlows** - Bioinformatics Workflows (BioPlatforms Australia; Australian Bioinformatics Facility)
An easy-to-use, web-based tool for bioinformatics workflows run on compute clusters and grids, with instances deployed at IVEC, VLSCI and QFAB.
- **Biosecurity Collaboration Platform** (Australian Animal Health Laboratory; national Consultative Committee on Emergency Animal Disease)
High-resolution visual telecollaboration platform for sharing biosecurity data and information.
- **DataMINX** - Data in Microscopy, Imaging, Neutron, X-ray (NCRIS Characterisation)
Data and metadata management for characterization facilities.
The DataMINX project was subsequently split into three projects focused on different facilities: the AMMRF (the NeAT PCA project), and the Australian Synchrotron and ANSTO (ANDS Data Capture projects).
- **DIAS-B** - Data Integration and Annotation Services in Biodiversity (Atlas of Living Australia)
Metadata repository supporting registration, discovery and search for biodiversity information; tools to support user annotations, e.g. for flagging problems with data quality.
- **Human Variome** - Software and Data Support for the Australian Node of the Human Variome Project (Human Variome Project; Molecular Genetics Society of Australasia; genetic disease clinics and researchers)
Searchable repository of genetic test results and interpretations collected from multiple pathology laboratories.
- **MACDDAP** - Marine and Climate Data Discovery and Access Project (Integrated Marine Observing System; Australian Ocean Data Network)
Enhancements to improve functionality and add support for Open Geospatial Consortium (OGC) and metadata standards to several tools, data servers, metadata harvesters and portals for the marine and climate science communities.
- **SISS** - Spatial Information Services Stack (Auscope; state Geological Surveys; Geoscience Australia)
Enabling a national geospatial data commons through the implementation of an open source software stack based on OGC standards, with an initial focus on supporting the requirements of the geoscience community.
- **NCJRND** - National Criminal Justice Research Data Network (Australian Institute of Criminology; state and regional crime research organisations)
Providing uniform access to criminal justice data sets held by various state organisations.
- **PCA**- Platforms for Collaboration in the AMMRF (Australian Microscopy and Microanalysis Research Facility)
Web-based tool for mapping research requirements to AMMRF facility capabilities; and data and metadata management services for users of instruments at AMMRF facilities.
- **PODD** - Phenomics Ontology-Driven Data management (Australian Plant Phenomics Facility; Australian Phenomics Network)
Repository providing flexible management of disparate data sets for plant and mouse phenomics, using community ontologies for describing the data.
- **Remote CT** - Remote Computed Tomography Reconstruction, Simulation and Visualisation (Australian Synchrotron; ANU micro-CT facility)
Remote execution of computed tomography calculations and 3D visualization for large-scale CT data sets.

OUTCOMES OF NEAT PROJECTS

NeAT projects developed or significantly improved over 30 software tools, all freely available as open source software. All projects deployed eResearch services based on these tools, in most cases across multiple organisations. Hundreds of research data sets are being made accessible (or more easily accessible) from these services. While each project was aimed at meeting the needs of a particular research community, in many cases the tools developed are more broadly applicable, and some are already being used in other discipline areas or by other NeAT projects.

The NeAT program aimed to increase the number of people in the sector who have the skills and experience to develop or use eResearch tools. NeAT projects have fully or partly funded over 100 people across 20 institutions, with many others from more than 50 organisations contributing to the projects as part of the co-investment of resources from project stakeholders. Hundreds of researchers from many research communities were involved in the implementation and testing of the tools and services deployed by the projects, and many more are now using them to enhance their research activities. The projects also raised awareness of eResearch and PfC activities in these research communities.

The NeAT projects provide exemplars of how research practices can be improved or transformed by the use of eResearch tools. Researchers and research organisations are reporting that the NeAT-funded tools are having significant impact on their research or on how they deliver data or eResearch services to their research community. With many of the NeAT services having only recently moved to production deployment (particularly for the second round NeAT projects), usage and impact will ramp up significantly over the next year. All of the tools and services provided by each project will continue to be supported by the project partners in the foreseeable future.

ABOUT THE AUTHOR

Dr Paul Coddington is Deputy Director of eResearch SA, where he has worked on eResearch projects and infrastructure since 2002. From 2007 to 2011 he also worked for the Australian Research Collaboration Service (ARCS), primarily as manager of the National eResearch Architecture Taskforce (NeAT) program, which funded projects to implement eResearch tools and services for many national research communities. He has undergraduate degrees in physics and mathematics from the University of Western Australia, and a PhD in computational physics from the University of Southampton, where he programmed some of the earliest parallel computers. He has 25 years of experience in eResearch, having worked at Caltech, Syracuse University and the University of Adelaide on a variety of research and development projects focusing on the application of high-performance and distributed computing and the Web to a variety of scientific problems, including the development of online scientific data repositories.