

# Brief Review of eResearch Service Provision in CSIRO 2019

*Justin Baker*

Justin Baker<sup>1</sup>, John Zic<sup>2</sup>

<sup>1</sup>CSIRO, Clayton, Australia, [justin.baker@csiro.au](mailto:justin.baker@csiro.au)

<sup>2</sup>CSIRO, Sydney, Australia, [john.zic@csiro.au](mailto:john.zic@csiro.au)

## BACKGROUND

CSIRO has advocated adapting science and research practices to deal with the large, complex data sets along with simulation and modelling carried out on advanced computational infrastructure [1]. It continues to develop its practices to ensure high impact science outcomes through its Data Schools and more recently, the Digital Academy as well as larger scale CSIRO initiatives such as the Managed Data Ecosystem and Digital Transformation which are currently in development.

The transformational aspects of adopting eResearch practices meant that there was, for many years, a piecemeal and fragmented approach taken by individual researchers, not only within CSIRO but also world-wide [2]. A social change is required in the way that science and research is carried out. The volume, availability and complexity of data, as well as the availability of new computing facilities and advanced computational science services means that business as usual is no longer possible.

An initial eResearch support service within CSIRO started in 2011 with the introduction of "Accelerated Computing Projects" (ACPs). The focus of ACPs was porting software to multicore/GPU/HPC environments, performance optimisation and numerical error analysis. ACPs were entirely funded by CSIRO's Scientific Computing program rather than individual research groups [3].

The original ACPs developed into "eResearch Collaboration Projects" (eRCPs) which have grown in scope with the addition of expertise that now includes Machine Learning, Data Analytics, Scientific Visualisation, Workflow Management and Science Data Handling.

The eRCP process is as follows:

1. Every six months, a "Call for Submissions" is sent out to all researchers across CSIRO. Interested staff are encouraged to submit a brief project proposal.
2. The researcher seeks formal endorsement of their project proposal by their program director.
3. After endorsement, the researcher submits the proposal for consideration to Scientific Computing.
4. After the Call's closing date, all endorsed submissions are reviewed and prioritised by members of CSIRO's Scientific Computing program.

Specialists from the Scientific Computing program are then assigned to work on one or more approved eRCPs. Over the six-month cycle, the resource allocation is around 0.2 FTE, with each staff allocated 3 eRCP projects per cycle.

## TODAY

The eRCP process continues to be very successful and is run as a competitive grant process within CSIRO.

In the most recent cycle, forty Scientific Computing Services specialists completed over sixty eRCPs from a total of eighty submissions. The underlying capabilities now include Performance Optimisation, Visual and Data Analytics, Science Data Handling, Software Sustainability and Science Workflows. The eRCP process also provides a mechanism to promote and introduce new tools to CSIRO's research community eg Jupyter and R/Shiny.

The submission and approval process has also been significantly streamlined. The new Scientific Computing eRCP portal includes semi-automated mechanisms for seeking endorsement along with directory name searching and linkages to related research projects. At the end of each cycle, a post-project survey is used to elicit feedback from individual researchers.

As expected, eRCP work has resulted in a range of useful outcomes such as a virtual reality tool to visualise biomedical images, an application to interactively compare multiple microbial genomes, and an improved workflow for 3D benthic habitat reconstruction.

## SOME ROOM FOR IMPROVEMENT

In spite of its ongoing success, the eRCP process is subject to review and continuous improvement. Some recent concerns/issues include:

1. Complexity of assigning IT staff to eRCPs:
  - The significant number of eRCP submissions received each cycle, combined with the large pool of scientific computing specialists, makes assigning staff to projects a non-trivial exercise. Other factors to consider include weighting organisational science priorities, servicing an appropriate mix of Business Units, as well as ensuring appropriately skilled software engineers are assigned to each project.
2. Lack of recognition:
  - Many eRCPs ultimately lead to the publication of scientific research papers. We believe scientific computing specialists who have made more than a minor research contribution via an eRCP should be included as contributing authors. More work needs to be done to ensure this occurs.
3. Occasional project failure:
  - eRCPs occasionally fail to meet their goals for a variety of reasons. These include, for example, poor scope definition or scope creep, limited client engagement and lack of access to existing code or data. To mitigate some of these issues, at eRCP acceptance, researchers are asked to provide access to all relevant data and to make themselves available when possible throughout the eRCP cycle.
4. Lack of skills
  - This is a general resourcing issue internationally, but lack of expertise in some specialist areas is making it difficult to fulfil some CSIRO researchers' eRCP requirements. This includes a need for machine learning, data analytics and HPC programming.

## TOMORROW

Future plans for the eRCP include:

- Process improvement and tracking
  - At all stages, from project proposal, endorsement, submission to project execution and close.
  - Reporting service introduction, such as a dashboard view of the entire set of project proposals.
- Communication of Scientific Computing's capabilities through the use of:
  - *Capability Champions*. These dedicated specialists within the Scientific Computing Group have recently been tasked with enhancing their respective capability and providing a central point to assist researchers with the eRCP engagement. This might be done, for example, by fostering communities of practice or conducting training workshops. An additional incentive for the Capability Champions is that they may be able to draw on their experiences in moving to a technical leadership role.
  - *The Research Software Engineering movement*. This initiative started in the UK and is about achieving better recognition for software engineers working in academia. They are often talented individuals with highly specialised skills whose efforts go largely unseen outside of their own organisation. The RSE movement is a concerted effort to address that deficiency and should be actively embraced within CSIRO, specifically for software engineers working on eRCPs.
- Broadening the scope
  - The eRCPs are available only to CSIRO's internal research community. With some modification, the model could be expanded to include external clients such as Australia's other publicly funded research agencies.
- Dropping the "e-" prefix on "eResearch", "eScience" and "eResearch Collaboration Projects". These become just "Research", "Science" and "Research Collaboration Projects" advancing the ongoing adaption of scientific work to embrace the digital (at all levels, from researcher to organisation).

## REFERENCES

1. J. A. Taylor, J. Zic, and J. Morrissey, "Building CSIRO e-Research Capabilities," in *eResearch Australasia 2008*.
2. T. Hey, S. Tansley, and K. Tolle, "The Fourth Paradigm: Data-Intensive Scientific Discovery," *Data-Intensive Sci. Discov. Microsoft Res.*, 2009.
3. S. Moskwa, "The Accelerated Computing Initiative," in *eResearch Australasia, 2012*.