Australian Government

NCRIS
National Research Infrastructure for Australia

Status Report on the NCRIS eResearch Capability

Summary

A Report to the Australian Government Department of Education and Training

Status Report on the Australian National Collaborative Research Infrastructure Strategy (NCRIS) eResearch Capability—Summary

Emeritus Professor Tom Cochrane
March 2015

A report to the Australian Government Department of Education and Training

For enquiries please contact NCRIS@education.gov.au

CONTENTS

1. INTRODUCTION

This is a summary report of the full Status Report on the NCRIS eResearch Capability, provided to the Australian Government Department of Education and Training in January 2015. The Summary has been prepared in the Department based on the full Status Report submitted by Professor Cochrane.

Preface

The period 2006 to 2014 has seen an approach to the national support of eResearch infrastructure by the Australian Government which is unprecedented. Not only has investment been at a significantly greater scale than previously, but the intent and approach has been highly innovative, shaped by a strategic approach to research support in which the critical element, the catchword, has been collaboration.

The innovative directions shaped by this strategy, under the banner of the Australian Government's National Collaborative Research Infrastructure Strategy (NCRIS), have led to significant and creative initiatives and activity, seminal to new research and fields of discovery.

Origin

This document is a Technical Report on the Status of the NCRIS *eResearch* Capability. It was commissioned by the Australian Government Department of Education and Training in the second half of 2014 to examine a range of questions and issues concerning the development of this infrastructure over the period 2006-2014. The infrastructure has been built and implemented over this period following investments made by the Australian Government amounting to over $430 million, under a number of funding initiatives.

Remarks on context for the development of this Report

The period covered by this Report has seen extraordinary growth and development in eResearch infrastructure in Australia. The investment has been catalytic for much research activity in Australia. Innovative and in many ways world leading, this approach to infrastructure investment can well be described as transformative across many fields, and despite some partly anticipated administrative complexity, its collaborative intent has been clearly realised in implementation and practice to a degree unmatched (in some respects) in other developed countries.

But the development of this Report has occurred in the context of a perceived uncertainty in the future approach to eResearch infrastructure. Naturally many of the responding parties are aware of this uncertainty, and it is likely that the commentary about sustainability of the approach, and funding cycles, has been perhaps more emphatic than would have been the case in an environment of greater assurance.

It has been a characteristic of research infrastructure funding, particularly that associated with eResearch development, to be cyclical, and to be framed in a future context which can be unclear. Australian research infrastructure funding shares this characteristic to some degree with similar activity in other jurisdictions.

There is a closely coupled view evident in the input to this Report that a more integrated approach to building capability may have been achievable, perhaps through a more durable leadership charged with ongoing oversight. On the other hand, it can be seen that by nesting responsibility for collaborative infrastructure development within the sector rather than over it, new collaborative responses and behaviours were encouraged and indeed resulted, as testified by those within the sector. The alternative, to vest significant sums with third parties, companies charged with pursuing these outcomes, was not considered to be then viable, based on all relevant experience to date at that time.

2. METHODOLOGY

Approach and Method

This report was developed by an Independent Reviewer, Professor Tom Cochrane, assisted by a Review Panel. The principal element of the work has been to consider the investments to ascertain:

- How relevant and critical they are in supporting Australian researchers.
- The relationship between these investments and other research infrastructure and research investments.
- How the capabilities and expectations of users have evolved and been addressed over time.

This has included a review of governance mechanisms to evaluate whether the pathway for coordination and implementation has been the most appropriate, the rate of evolution of eResearch infrastructure broadly and external influences that affected the development of the infrastructure.

The Terms of Reference for the Report's development are at **Appendix One**.

The stakeholder group consulted comprises organisations using the provider services, these being research agencies and institutes, universities, research centres, NCRIS domains with high use or dependency on the eResearch Capability, the Australian Academies, the ARC and NHMRC, state and nationally based e-infrastructure providers, some selected vendors (to the Capability). It also includes individuals involved with planning and implementation processes, particularly as established in groups to assist and advise government consideration of the Capability over the last several years.

The matters considered were examined through views solicited through written comment on questions based on the Terms of Reference, or through consultations. A Call for Submissions was made in mid-August 2014, with responses received in September and October. Consultation has occurred through the period from late August to December 2014. Some respondents contributed both a written submission and attended a consultation. Total written submissions, including the seven NCRIS eResearch service providers, numbered 60.

A list of respondents and consultations is provided at **Appendix Two**. A summary of the outcomes of those consultations with stakeholders responding to the Terms of Reference, and which form the principal basis of this report's findings and recommendations, can be found in Section 3 of the full Status Report prepared by Professor Cochrane. An abridged version of that summary is at **Appendix Three** in this report.

Following the analysis and writing up of the submission content, further research community inputs from the responding organisations and individuals have been gained in a Workshop session held in Canberra in early December 2014, described in *Section 5*, Further Analysis.

References used in the development of this report are provided at **Appendix Four**.

3. FINDINGS

Overall Findings

Quantum and Approach

The Australian eResearch Capability has been built up through an investment level that is very significant by any reasonable relative standards. The national eResearch investments have been timely and critical to the advance and competitiveness of Australian research. They have been engineered to gain significant co-investment aligned with Australian (state) government and institutional (universities, research agencies and institutes) resourcing. The deliberate strategy of fostering and implementing collaborative approaches has been a hallmark of the development as acknowledged across a range of community responses.

Within that recognition, there could nonetheless have been better national coordination, and more focus on skills and expertise, and in some cases better execution. Within the research community there is a variety of views about the means of deploying the Capability and the balance and conceptualisation of the investments, which are elaborated in this Report.

Evidence

Throughout this Report data and statements in support of views about the effect of the investments can be found. This richly complements other material reported to the Department during 2014.

There is a particular issue however about the partnership with industry in its use of the infrastructure. Community responses that may have elicited more evidence on this are somewhat sparse, and this is part of a wider picture in Australia about the connectedness between industry use of and benefit from research infrastructure.

Forming conclusions on this can be fraught. On the one hand, any reader of the *Tildesley Report* in the UK cannot help but be struck by the more evident industry/academic partnership, and this is underscored by the fact that Tildesley himself has been both an industry leader and academic. On the other hand, there is some history in Australia of links between industry and research being in place but less visible, indeed there is some reasonable conjecture that this opacity is in part due to a disinclination to report the linkages. Some robust debates in the not too distant past about the merits of seeking to measure "research impact" (cp "quality") are a related issue.

Dependencies on the Capability – Research and Operational

It is evident from the substantial feedback received, that considerable areas of Australian research activity are completely dependent on the services and assets provided by the Capability. This in one sense could be taken as read, but it is clear that voice has given to this dependency as a means of expressing insecurity about ongoing support for it.

A particular issue warranting focus and reflection is the extent to which the transformations that have occurred, particularly in the last three to five years, have developed new operational dependencies on the way data is managed and science is done. Enormous benefits have resulted from the deployment of the infrastructure and for some the issue of sustaining the new ways of doing things as their ongoing and essential business has become a vital issue.

Findings against the Terms of Reference

Appropriateness – Market Failure

The view that the Australian government investments are a response to market failure is expressed by almost all respondents and providers.

In some cases some elements of the Capability are differentiated in the discussion of responding to market failure. Thus federated access and approaches to research data storage may have developed in some other way. Frequently views have included comment on the rate and likely timescale for any similar development in the absence of Australian government investment. Some commentary also includes speculation about the price differences if the needs had been filled by commercial offerings. At the same time, commercial offerings are also the subject of comment about new developments towards the end of the nine year period of the investments.

Appropriateness – Transforming Research

There is a widely held view that the effect of the investment has been transformative.

In some cases this transformation will be more evident over time, and there has clearly been a more dynamic impact in some disciplines than others. But there are also descriptions of more general impact, including the notion that the deliberate construction of infrastructure investments as "collaborative", generated further (and desirable), collaborative behaviour among researchers. There seems to have been a cultural effect of the investments on institutions and disciplines themselves, including the collaboration of research groups with each other, research groups with support groups, and among support groups.

Appropriateness – Meeting Expectations

For many, expectations have been met, in some cases exceeded, and in others, clearly not met. This varies by the elements of the Capability, though not consistently as perceived by different stakeholders. Broadly similar findings are evident in the reflections on stakeholder expectations, with some providers offering comment on work done to gauge stakeholder satisfaction.

Contribution to Technology Platform – as a Whole, or in Selected Disciplines

There is a significant view that the investments have led to the development of such a "platform".

There is however, some disagreement here, including around the issue of what constitutes a "platform", as well as discussion of the way data has been the object of specific investments. This includes some tension between views about data as an asset in "e-infrastructure", or investment in research data more generally as assistance to research infrastructure development in the nation.

The somewhat related question of the proportionality of investment in third and fourth paradigm eResearch approaches/facilities also arose in some of the discussion inputs, but the distinction is probably of limited usefulness[1].

Contribution to Platform – Complementarity with Investments at Other Levels

There have been widespread complementary investments at state and/or institutional level. There is significant commentary about the coordination of these investments, and the extent to which they may have realised greater potential if better coordinated.

[1] The term "fourth paradigm" has been deployed after the 2009 publication of the title of the same name by Microsoft, to attempt to describe and differentiate data-intensive science as a new phase of research, along an evolution of eResearch beyond computationally based modelling and simulation, ie "third paradigm".

Contribution to Platform – Future Leveraging

Almost all responses describe a future intention to leverage, although the question seems to have been interpreted as also applying to the leveraging of future investments as well as those to date.

Governance

There are mixed views about governance, ranging from satisfaction (and some praise), to dissatisfaction. This issue in particular sees stakeholders differentiating elements of the Capability.

The most general theme emerging is widespread concern about national coordination and coherence of the elements of the Capability[2].

Take Up – Extent

The extent of take-up has been described as generally good, with measures given by providers. Concern about two of the elements developing more slowly than expected is found among many stakeholders.

Take Up – Improvement possibilities

There are ways in which take- up might have been, and could in future be improved.

These include more focused effort on outreach and awareness raising, stronger coordination among the elements, more technical support skills, a focus on relevant training for research groups and the need for greater funding certainty.

Take Up – Barriers

The barriers nominated include wide concern about investment in skills and expertise, as well as the recurring theme of the need for stronger integration, and more sustainable funding. Providers mentioned the complexity and novelty of their service offerings in some cases, as a barrier.

Lessons Learned

Themes here include the need for stronger focus on:

- communication and marketing
- integration, including addressing the way in which engagement has been complicated by the separate branding of the elements of the Capability
- project management and governance
- codifying practical learnings from the experience of different rollouts for different elements, ways of engaging end-user researchers more systematically in the development of the investment plans
- addressing tension about the overall approach in the distribution of funds among the various "horizontal" components.

Providers also identified better alignment and coordination, and the challenge of carrying out complex procurement to deadlines as areas where lessons could be learned for the future. In addition, there was some reflection on the extent to which investment might have been better if lead by domain or discipline groups rather than functionally, however, this was not a majority expressed view.

Comparisons - International

The general finding is that there is significant confidence and pride in the achievements in Australia to date, particularly in terms of innovative national approaches in data and federated access.

[2] It should be noted that governance and other aspects of the RDSI element were the subject of separate review in late 2013/early 2014, and the project's governance structure has subsequently been revised following agreement between the Department and the Lead Agent.

In the community commentary, the elements most often singled out as having performed very well by international standards are the Australian National data Service (ANDS) and the Australian Access Federation (AAF), both cases where the Australian approach has been world leading. There is also significant satisfaction with the maturity and performance of peak facilities, the computational infrastructure at the National Computational Infrastructure (NCI) and the Pawsey Supercomputing Centre.

However, there are some expressions of concern within this overall sense of confidence. These involve:

- the sustainability of the approaches taken in Australia
- the extent to which the HASS disciplines have not developed at the same rate as international comparators
- Australia's overall level of investment in HPC Capability at both peak and institutional level
- the integration of the effort.

Providers themselves offer a mixed commentary on the way they compare, referencing in some cases the innovation shown in Australia compared to elsewhere.

Comparisons - Other Industry

There are mixed and ambiguous views about the comparison between Australian eResearch infrastructure investments and development in other industries or sectors. Relevant here also is the discussion about the opacity of the industry/infrastructure nexus in Australia.

Influences and Constraints - Budgetary

There is widespread concern about the sustainability of the budgetary approach. This includes reference to the extent to which innovation might be restrained at the local level because of a lack of confidence about resource replacement after a granting or investment phase is completed.

Also attracting widespread comment are well-known constraints relating to capital versus operating expenditures which were the conditions of some of the investments. It is worth noting that by and large, there is no criticism overall of quantum, but there is commentary on balance, particularly insufficiency at peak and university levels for high-end computing.

Influences and Constraints - Structural

The structural issues identified revolve around fragmentation in governance, the unwieldy nature of managing state level co-investment, and the insularity of projects as they have developed.

Influences and Constraints - Technological

Three providers started operations in an environment where the technologies applicable to a significant part or the whole of their missions were still maturing, and this has arguably had an impeding effect. Although some of the investments led to projects that were themselves experimental, technology "deficits", have not been significantly identified by the stakeholder community, with the one significant exception, namely, skills and expertise in human resources.

4. CONCLUSIONS AND RECOMMENDATIONS

Noting that the consideration of future strategy and approach is a separate task, there are nevertheless conclusions that can be drawn from the findings outlined in this Status Report, and recommendations which arise from those conclusions. These recommendations take the form of matters to be considered in future planning: depending on the individual suggestions these may be taken up by the Department directly at an appropriate time, and/or referred to a future eResearch framework development process.

These six conclusions and their attendant recommendations are made in the overall context of ***two high level conclusions***, which are that:

1) The NCRIS approach has succeeded in providing significant eResearch infrastructure benefits for research in a way that has accelerated and enhanced collaborative activity among both infrastructure providers and researchers
2) While other possible national investment of a complementary nature—and in particular to catalyse skills development—would have been of significant benefit, the quantum and proportional allocations to the elements of the eResearch Capability have been broadly appropriate.

I) Research Community and Funding Council Linkages

The consultative process adopted in Australia in the establishment of the NCRIS eResearch program has been thorough and commendable. However, there is a need to ensure the process of research community consultation in shaping the investments is extended, and continued. In particular, there is a perceived disconnect between the infrastructure decision making process and the planning and operations of the funding Councils, the ARC and the NHMRC.

Recommendation: Future eResearch infrastructure strategy include the consideration of the means by which research community inputs and funding Council linkages will be developed, and maintained.

II) Strengthening of Governance

The governance practices and standards have been significantly variable across the elements of the NCRIS eResearch infrastructure.

Recommendation: In any future deployment of a Lead Agent based investment approach, specific and detailed project standards be in place for governance and management.
These standards will need detailed development, and should specify the ways that Lead Agents and Boards: a) provide oversight; and b) guarantee optimal coordination and national integration.
This approach should be enhanced through regular intersection with a small coordination and leadership group whose purview crosses eResearch projects, and which provides advice and cross-project integration.

III) Flexibility in Facilities and Service Development

The national eResearch infrastructure should have many functions and services which are commercially sourced. It is also the case that project framing, funding, and implementation is often on a timescale slower than, and inconsistent with, rapid evolution in the marketplace.

Recommendation: Greater, more explicit flexibility should be built in to future project approaches to allow the making of fit-for-purpose choices at relevant points in project development, with particular reference to commodity procurement to underpin service provision.

Constraints which would oblige outlays on capital rather than operating expenses, (or any similar limitation) should be eliminated, or minimised.

IV) Skills and Expertise

There is continuing widespread concern about skills and expertise deficits. This concern has persisted throughout the life of the NCRIS investments, and involves both research workforce issues and technical skills in supporting eResearch infrastructure. Significant recommendations for action at the beginning of the NCRIS funding period remain to be addressed, and are generally as pertinent in 2015 as 2006.

Recommendation: Future framework planning consider the continuing need to address both research workforce and technical support skills and expertise, and take up the issue of preferred strategies at both institutional and national levels.
This should involve research agencies and institutions.
Discussion of the approach with the university sector collectively is needed.

V) Approach to the Provision and Dispersal of Funds for Infrastructure

The collaborative intent of the NCRIS approach can be frustrated by the adoption of competitive processes for funding allocations.

Recommendation: Where a Project is charged with a further subdivision and allocation of funds to develop infrastructure, the potentially inhibiting effect of any competitive allocation process on collaborative approaches be anticipated, and the risk mitigated.
Accountability for the decisions taken in such allocation processes should lie beyond the Project itself.

VI) Industry Involvement

While there is a history of linkages between industry and government in the development of research policy, the connectedness between Australian industry, broadly defined, and the eResearch infrastructure could be improved.

Recommendation: Consideration be given to a more formal involvement by industry in framing future national approaches to Australian eResearch infrastructure.

5. FURTHER ANALYSIS

As noted, the expert and community input to the development of this Status Report led to the further identification of issues and themes, the discussion of which was the basis of the 5 December Workshop. This activity, open to all responding parties, saw discussion framed around seven significant themes, which are detailed at Appendix Four of the Full Report. The further analysis summarised in this section emerged from the discussion of these themes.

The Model (Structure and Coordination)

Widespread observation about the model for the NCRIS investments was a feature of the community input into the Report. In particular, commentary about the Lead Agent model[3] employed by NCRIS projects was diverse and extensive. The significant convergence of views was that better coordination must be achieved among the facilities and services that have resulted from the investment period 2006 to 2014 and, as a consequence, the ongoing question of the leadership of that coordination—both institutionally and nationally—is paramount. It is not a finding of this Status Report that the Lead Agent model did not achieve substantial collaborative benefit. This approach is therefore, not one to be abandoned *in toto*.

It should be noted that the *2010 NCRIS Evaluation Report*[4] outlined a number of key learnings, but did not have as a focus the performance of the Lead Agent model *per se*. Subsequent to that *Report*, the flexibility of the lead agent model, and the evidence and future likelihood that it may encourage co-investment were identified in the *2011 Roadmap* (p 14).

The Australian approach has also attracted international attention. In work conducted in Canada on *Advancing Canada's digital infrastructure*, the approach taken in Australia, particularly the identification of critical enabling capabilities under the rubric of eResearch, is described, with particular reference to "transparent national consultative processes" that were adopted. This (Canadian) report goes on to comment about problems "inherent in asking individual institutions to take the lead role in managing major national programs" (pp 17-18), however, the report does not engage in further discussion of either the model, or alternatives to it.

One view was that, although there has been some controversy around the Lead Agent model, it has to be acknowledged that a higher degree of collaboration and consultation within the research sector, along with expectations about the outcomes from that, has been a feature of the NCRIS investments. While the view that universities were not the best option for the management of national infrastructure and service development and delivery was strongly expressed, there was a counter view that comparisons in other jurisdictions (such as the US and the UK) lead one to draw conclusions about other forms of fragmentation. There was acknowledgement that the Lead Agent model would be more effective if there was greater clarity about governance, expectations, and coordination. The *2014 NCRIS Survey* (not published) responses tend to reinforce a view of divergent approaches among Projects (elements) to governance, communications and marketing. It is worth noting that the topic of governance of "e-infrastructure" is gaining greater focus internationally, particularly as research support moves from "project style provision" to larger scale "production quality", where the transition amounts to a

[3] Whereby both NCRIS funding for, and responsibility to the relevant Commonwealth Department for the delivery of, an NCRIS facility or network is vested in a single Lead Agent via a Funding Agreement with that Department. That facility or network is, by virtue of being part of NCRIS, an open national, collaborative infrastructure facility, or network however, which the single Lead Agent must deliver. All facility/network arrangements for governance, co-investment, access and sub-contracting sit below the Funding Agreement with the Department, and are ultimately the responsibility of the Lead Agent.

[4] Note that full references are provided at **Appendix Four**, to all works listed here in italics.

significant change in the governance of those research support activities (*The Emerging Governance of E-Infrastructure, p 114*).

During the Workshop discussion of the model, the issue of "reach" was identified, noting that the model has served communities which were "ready to go" well, and has not perhaps pulled in a wider range of more diverse domains. This is a significant issue, as policy developed in the past, and enacted through the Australian approach, has been focused on reaching a wide number of researchers, and lifting the country's research performance across a wide spectrum.

Collaboration

The deliberate choice of a collaborative rather than competitive approach to infrastructure planning has been a hallmark of the period under review, and was reiterated in the *2011 Roadmap*, (p.9,12). Furthermore, the intent has been to meet the needs of "a broad base of users", as well as to support novel approaches and innovation.

Collaboration in the provision of infrastructure has been a hot topic of recent mention in other jurisdictions (Canada, Europe, UK et al). In the recent evaluation of projects on the *ESFRI Roadmap*, the high-level expert group emphasised the need for greater coordination and leadership:

> *participating institutions could have developed a joint work-plan, funding plans, governance and project organisation for a fully integrated distributed research infrastructure, they showed insufficient inclination to do so and the outcome presents significant gaps* (p 4).

The Canberra workshop reaffirmed the examples of significant collaboration among infrastructure services and facilities in Australia. Observations at the workshop included the view that collaboration has led to greater coherence across the research system, making Australian researchers more competitive, with numerous examples cited of collaboration across various elements of the infrastructure (eg NeCTAR Virtual Laboratories, the Research Data Storage Infrastructure (RDSI), ANDS). It was acknowledged that mechanisms generated by the NCRIS approach had brought communities together, obliging them to collaborate in order to secure facilities and services – although the view was also expressed that what was described as "forced collaboration" could develop situations where it had not worked well.

The discussion also centred on various vehicles for collaboration, noting that competition and collaboration are not necessarily mutually exclusive processes, and that in some cases it may be desirable to engender friction leading to change and growth.. It was noted that industry has many examples of collaboration on joint facilities or infrastructure, occurring among fiercely competitive entities. By and large, universities have not developed evidence of this level of maturity regarding collaboration within competitive environments, although it was noted that at the discipline level, for example astronomy, aggressive competition which in turn relies on close collaboration for research infrastructure, has some similarity to the industry example.

As in the submissions to this Report, the view was expressed at the Workshop that shared research needs act to engender collaborative relationships among researchers, but the extent to which this has occurred is dependent on the relevant discipline.

Domains, Platforms and Disciplines

The approach in Australia, encapsulated by the NCRIS program, has been to support research through the provision of research infrastructure which is in part domain aligned, but which is in turn underpinned by collaborative capabilities in a "layered functionality" (*AeRIC Annual Report*), which the eResearch investments typify. As stated in the *2011 Roadmap* (p 10), a "driving principle… was that the research infrastructure… should support a large number of researchers, often across a range of disciplines."

The workshop discussion made clear that while some domains can, in theory, organise themselves, not all communities are well represented in the Capability's users. It was noted that the way investment is

allocated and planned may engage smaller communities if done in a particular way (for example developing discipline oriented Virtual Laboratories), but affect much wider groups when organised as national programs – such as ANDS.

Differences applying to the Humanities and Social Sciences were noted, with a view that collaborative use of eResearch infrastructure by scholars in these disciplines is relatively recent. Differences among disciplines in terms of "readiness" to use eResearch infrastructure are reasonably well understood, however, there remains a general tension in seeking to develop the research infrastructure between "push" and "pull" tendencies[5].

The workshop delivered something of a consensus that the current balance between investing in platforms and domains is about right, but that there is a more significant issue in ensuring balance within the domains, including their ability to be heard during planning and investment setting processes. A concluding suggestion was that the needs of domains in the context of a broader "coarse-grained" approach to the research capability investment be more specifically addressed in planning.

While the question of the extent to which the disciplines identified around groupings in the *2011 Roadmap* have been served by the NCRIS investments was not conclusively discussed at the Workshop, the concerns expressed by the HASS community about the difficulties inherent in requiring many fields to coalesce around a single theme[6] were made clear.

Central to this suite of issues is the matter of maintaining research community connectedness, which is the subject of the first recommendation in this Report.

Data and Data Centric Science

Data and its treatment and role in research, is a vast topic which has received separate and detailed consideration in work recently commissioned by the Government, *The Australian Research Data Infrastructure Strategy,* (preceded by the discussion of data as a priority in the 2011 Roadmap, pp 77 and 78, and other places before that, eg the *eResearch Coordinating Committee Final Report* – 2006). None of the 18 recommendations resulting from this work (ie *The Australian Research Data Infrastructure Strategy*) is contradicted by inputs to this Status Report, and indeed those concerned with changing work practices (eg Recommendations 15, 16, 17), have been affirmed in the context of the need for greater focus on skills, expertise and leadership.

Another apparent tension about data in the national landscape concerns the "peak" versus "long tail" issue. This point concerns an observation that national eResearch investment should focus on data at large scale and/or of significant complexity, and not replace appropriate levels of institutional policy attention and investment. The view that there should be less or no national investment to improve practice relating to data storage management and deployment that might be described as in "the long tail", was not supported by the Workshop discussion.

One point that was reiterated in discussion concerned the fact that some of the national achievements through ANDS give Australian researchers and their institutions a more competitive position in a global context where the challenges of data at scale and the complexities associated with that are set to grow continually into the future. The workshop discussion raised several other points, including the future growth of data demand in the HASS disciplines, supported by the continuing unmet demand for digitisation; privacy and ethical issues relating to hosting and future availability of data; more useful ways of integrating data from different sources; the value associated with data "in use" as distinct from data

[5] Where "push" characterises the notion of technical experts with understanding and foresight spelling out what technology might do for a discipline, while "pull" characterises research groups and their leaders identifying what they want, to make progress in their discipline.

[6] In the case of the HASS disciplines, the Cultures and Communities Capability in the *2011 Roadmap*.

"as an asset", and the likely future need for a better and stronger focus on "data-using research infrastructure" compared to "data-collecting research infrastructure".

The openness and availability for reuse, of research data is also an important plank in ensuring greater research integrity. In this vein it should be noted that the rising importance of data in numerous ways gives rise to legal questions as well, with increasing demands for provenance tracking, and likely future emphasis on reducing the contestability of data sources with better management of linking and locating tools.

Commercial Offerings or Research Sector-Developed Infrastructure?

There is now a long established pattern, in which today's innovative and expensive high-performance infrastructure becomes tomorrow's lower cost, commercially available, "taken for granted" infrastructure. It is worth observing that whether publicly purposed national funds are spent on facilities developed within the "research sector", or offerings provided commercially, a significant proportion of the investments is taken up in the purchase of facilities and services from industry, i.e. the private sector. Indeed, in principle, the only significant continuing public investment which is deployed substantially in the public domain, is in the skills and expertise required within the research sector, about which nevertheless, there is continuing concern regarding insufficiency.

The Workshop noted that leading edge high performance computing, and cognate areas of the infrastructure are not profitable, and do not lend themselves to commercial provision because of their very nature. However, there is a generally supported view that if the national research infrastructure is viewed holistically it could and should have many elements which are commercially provided. Indeed the *2011 Roadmap* anticipated the role of commercial cloud service provision, "where appropriate", (p 79).

It was generally observed that because of dramatically falling prices for some aspects of commodity, and the continuing interest by commercial providers in responding to this market where it makes economic sense, the future will see a greater proportion of viable direct commercial services underpinning the provision of specialised eResearch services. (Of interest was the prevalence of the theme of commodity provisioning – infrastructure as a service – at the recent *eResearch Australasia 2014* conference). There was consensus that these issues should not be viewed ideologically or dogmatically. The national eResearch infrastructure will continue to be a mix of public leading-edge investment, and commercially provided facilities and services with the latter playing an increasing role.

From this discussion also came the point that project framing, funding, and implementation is often on a timescale slower than, and inconsistent with, rapid evolution in the marketplace. This leads to the view that greater flexibility should be built in to future Project approaches to allow the making of the best choices at relevant points in project development, and is reflected in Recommendation III.

It was also noted that on the issue of data in some parts of the national infrastructure, sovereignty will continue to influence choices made about commercial party hosting.

Industry/Infrastructure Nexus

The *2011 Roadmap* identified increased collaboration "between industry and research organisations" as an important policy issue, nominating improved access by industry to research infrastructure as a mechanism supporting the goal of translating "research outcomes into increased productivity" (p. 16).

However, it is clear that from an e-Research infrastructure provider perspective, as well as from the commentary of many of the stakeholders, that such improved access is unclear, not well documented, and, as noted at the Workshop, not facilitated by the current structures for the funding, development, and operational settings for the provision of Australian eResearch infrastructure.

There are some complex issues for consideration here. While it is true that there are unfavourable comparisons between Australia and other developed countries in terms of the industry/research nexus, the issue of the availability of research infrastructure within this as a separate issue in its own right is not a matter that has received particular focus. To some extent there is an inherent opacity in the way that industry engages with research *infrastructure* for the solution of problems. A frontline researcher engaged in an industry consultancy does not necessarily report to industry their reliance on institutional or nationally provided infrastructure in carrying out the work. From time to time audits of connectedness between eResearch infrastructure and industry have been carried out at state level, (eg Queensland, by Queensland Cyber Infrastructure Foundation (QCIF)), but have not led to persistent and durable statistical tracking.

In fact, the centrality of eResearch infrastructure in Australia to applied outcomes in areas such as land use and water management, indicates that it is now an imperative to understand that the infrastructure should be the basis for operational centres for various forms of sophisticated applied problem-solving, directly connected to Australian economic planning and development.

While there could well be a certain scepticism within Australia of simply adopting the slogan that to "out-compute is to out-compete", which is referenced in both UK and US thinking, it is also clear that (building on earlier patterns of defining research priorities, as in the *2011 Roadmap*) the specific and unique requirements of Australian industry (with particular reference perhaps to agriculture and energy), do need stronger mapping to the research infrastructure investments that may occur in future. Accordingly, some means of improving the maturity of the industry/infrastructure nexus is desirable, and is the subject of Recommendation VI.

Leadership and Expertise

On the issue of balancing the apparent tension between the investments in "peak capability"[7] and more general eResearch infrastructure development[8]—the so-called majority-peak tension—the Workshop noted that there was no clear methodology for determining balance between the components. (Or if there has been, it is not widely understood). An alternative view was expressed that rather than have a fixed methodology, the preference should be for some top-down coordination which leads decisions on the balance of investment as required for particular areas. It was also noted that the way that "peak facilities" have been funded had led to some expressions of concern about the need for greater investment in skills and software within those facilities, (rather than a mainly "hardware" view about where spending should be).

The evidence gathered through stakeholder responses suggests that the way leadership is exercised within institutions in Australia is variable and inconsistent. In future planning, it will be useful to engage institutions, universities in particular, in a more focused discussion about the institutional role in consciously developing eResearch infrastructures, and their support, to an appropriate level within and among institutions, nationally and regionally. In the case of compute cycles for example there appears to be no explicit formulation of standards of provision. Rather the terrain is marked more by intermittent investment approaches which in turn reflect the way research infrastructure has been approached by institutions, through the ARC Linkage, Infrastructure, Equipment and Facilities (LIEF) scheme, or regional partnership, or institutional resources, and sometimes a mix of all these.

On the issues of skills and expertise, the Workshop saw an engaged discussion, reflecting the extent to which these issues were raised in the submission process. There can be no doubt that some of the NCRIS programs, for example ANDS, have catalysed skills development in particular ways, (eg in the case of ANDS, data management). But deficits in technical support skills, particularly the kind needed for the

[7] For example in the Australian context, petascale high performance computing (HPC) infrastructure.
[8] For example, widely accessible software- and platform-as-a-service offerings delivered through a cloud computing platform.

development of new software code, as well as the rewriting and "hardening" of existing code, present continuing challenges. In this context, it should be observed that some of the NCRIS facilities have thought long and hard about building and retaining skills, including the development of career pathways. The NCRIS eResearch providers, according to *2014 NCRIS Survey* data, while nominating skills shortages as a significant problem, vary in the rating of risk in terms of the way that the relevant provider will meet deliverables, with most nominating skills shortage at the "serious" or "moderate", risk levels. Other institutions and state providers have been even less sanguine about the ability to retain highly talented staff. It was noted at the Workshop that the retention of expertise is seriously affected by funding uncertainty.

The issue of skills and expertise in eResearch is a global one, and receives extensive commentary in all developed countries as they consider future policy for the provision of facilities and services. The international critique addresses both academic workforce skills, and technical support needs and deficits. Indeed this is an area where relationships between industry and academe are competitive, and both industry and research institutions identify a continuing need for the development and refreshing of computational and software engineering skills. There is a clear strong view that improving the overall standards of skills and expertise should be tackled both nationally and institutionally.

In the case of the UK, the *Tildesley Report* has expressed the need to "develop a young, and growing, community of researchers who are able to exploit current and future leading edge… computing to the fullest extent", (p 17), and evident concerns about academic workforce skills are at least as significant as the need for workforce development for the providers and developers of data and software services. Workforce development is also one of the significant topics identified by the National Science Foundation's consideration of the *Cyberinfrastructure Framework for 21st Century Science and Engineering* developed during 2012. The expression of concern about skills and expertise deficits has persisted throughout the life of the NCRIS investments. At the beginning of the investment period in 2006, the *eResearch Coordinating Committee Final Report* (p ix), laid significant emphasis on the need for skills acquisition, particularly through specified undergraduate and postgraduate programs and training, targeted at the broader Australian research community. The importance of it has been referenced in every significant document since, and the demand for stronger leadership, at both institutional and national level is unabated. This matter, particularly at institutional level is the subject of Recommendation IV.

APPENDICES

APPENDIX ONE

Terms of Reference for the Development of a Report on the Status of NCRIS eResearch Capability

1. Purpose of the Report Activity

This Activity aims to report on the status of national eResearch infrastructure in Australia, funded under the National Collaborative Research Infrastructure Strategy (NCRIS) program and related programs, as of year's end 2014.

It will take a strategic view of investments to ascertain how relevant and critical they are in supporting Australian researchers, the relationship between these investments and other research infrastructure and research investments, and how the capabilities and expectations of users have evolved and been addressed over time. This will include a review of Governance mechanisms that supported the investments to evaluate whether the pathway for coordination and implementation was the most appropriate in the circumstances, the rate of evolution of eResearch infrastructure broadly and external influences that affected the development of the infrastructure.

The findings of the Status Report will inform future consideration of research infrastructure investment and delivery, and in the short-term will be used as input to the 2014 efficiency review of NCRIS projects and facilities.

2. Timeframe

The activity is expected to begin in August 2014 and conclude by January 2015.

3. Terms of Reference

The Status Report will examine:

a. how appropriate the national investments in eResearch infrastructure have been, including:
   i. the degree to which they responded to market failure;
   ii. their transformative impacts on the conduct of research including fostering collaboration; and
   iii. how they delivered against expectations of stakeholders and those defined by the funders ;

b. how well national investments in eResearch infrastructure have contributed to a platform of advanced information and communication technology (ICT) to support the research sector and the degree to which they intersected with institutional or state-based investments;

c. the appropriateness of the governance arrangements for the implementation of the investments;

d. the level of uptake of the infrastructure and any barriers to that uptake;

e. lessons learnt or systemic issues to be addressed to guide future investments in delivery of eResearch infrastructure;

f. how the development of the infrastructure compares to similar developments in eResearch infrastructure elsewhere, for example internationally, or similar developments in ICT support for other sectors; and

g. any external influences that affected the development of the infrastructure –these may include budgetary, research sector structures and arrangements.

The Status Report **will not** aim to provide an individual evaluation of each project nor seek to examine matters to be considered under the NCRIS efficiency review.

4. Process to Develop the Report

a. The Department will appoint an independent reviewer and review panel, and will consult a range of stakeholders in seeking suitable candidates. The reviewer and panel members will have:
   i. No irresolvable conflict of interest;
   ii. A demonstrated understanding of the role of research infrastructure at the NCRIS scale, and the role of government and stakeholders in supporting research infrastructure;
   iii. A proven record in conducting or playing a major role in like activities, including but not limited to similar reviews; and

iv. Demonstrated strategic and analytical capabilities commensurate with the scale of this review, including the ability to develop key recommendations

b. The reviewer and the Department will agree on members of a review panel which will:

i. Comprise no more than three members

ii. Have expertise on:

- ☐ like reviews;
- ☐ the role of research or eResearch infrastructure to support research;
- ☐ Australian research delivery or support on a significant scale; and
- ☐ project governance and management.

iii. Significant standing with the research or research infrastructure community.

c. The reviewer will lead the development of the Status report, with the assistance of the panel, and will be responsible for delivering the Report and recommendations noted below.

d. The panel will advise and assist the reviewer, and participate as appropriate in discussions and activities throughout the process.

e. The review will be assisted by Department in the form of secretariat services.

5. Status Report and Outcomes

The Report will be informed by invited submissions and discussions with relevant stakeholders. This may include at a workshop with stakeholders; individual interviews in addition to the workshop, as needed. Stakeholders may include, the lead agent, the implementers of national eResearch infrastructure projects, relevant NCRIS projects and facilities, the Department of Education, funders, key organisations involved in delivery or use of infrastructure, and other impacted parties.

The panel will submit a report to the Department, which will include:

a. Arrangements including these terms of reference.

b. Membership of the panel, including declarations of members' level of interest.

c. Methodology including a list of documentation provided to the review panel.

d. Key criteria for assessment and summary of the findings against the terms of reference.

e. Details of findings against the terms of reference.

f. Summary of the emerging trends and environment against which the eResearch infrastructure will continue to operate, to the extent possible.

g. The set of recommendations and proposed actions to be considered by the Department, in priority order.

6. Key dates

Dates	**Milestone**
Week Beginning 4 August	Appointment of Consultant after consultation by Department
Week beginning 4 August	Establishment of panel after consultation between Consultant and Department
Announcement by week ending 15 August; Invitations by 28 August	Advice to stakeholders and invitation to provide submissions or participate in an interview
19 September	Stakeholder submissions due
w/e 26 September	Panel meet to consider submissions
By 10 October	Panel meets with stakeholders by workshop
W/e 17 October	Outline Document with preliminary view to Department
21 November	Interim report to the Department
TBA	Reviewer and potentially members of the Review Panel meet with the Department to discuss report and receive feedback
TBA	Reviewer finalises Review report and forwards to the Department
TBA	The Department advise all parties (as appropriate) of Review outcomes
TBA	The Department to implement Review recommendations as appropriate

APPENDIX TWO

List of Respondents and Consultations

A. Submissions, listed by Respondent

The following organisations or individuals made submissions. In some cases these were supplemented by consultations.

PROVIDERS (ie Elements of the NCRIS eResearch Capability)

Australian Access Federation (AAF)
Australian National Data Service (ANDS)
National Computational Infrastructure (NCI)
National eResearch Collaboration Tools and Resources (NeCTAR)
The National Research Network Project (NRN)
Pawsey Centre
Research Data Storage Infrastructure (Project) (RDSI)

STAKEHOLDERS (Including Lead Agents)

Group A: Research Agencies, Centres and the Academies
Academy of the Social Sciences in Australia (ASSA)
ARC Centre of Excellence for Climate System Science
ARC Centre of Excellence for Integrative Brain Function
ARC Centre of Excellence in Molecular Imaging
Association of Australian Medical Research Institutes (AAMRI)
Australian Academy of Science (AAS)
Australian Academy of the Humanities (AAH)
Australian Institute of Marine Science (AIMS)
Australian Nuclear Science and Technology Organisation (ANSTO)
Australian SKA Office
Commonwealth Scientific and Industrial Research Organisation (CSIRO)
Cooperative Research Centres Association
Geoscience Australia (GA)

Group B: NCRIS Domains and Collaborative eResearch Service Providers
AARNet Pty Ltd
Astronomy Australia Ltd
Atlas of Living Australia
Australian Urban Research Infrastructure Network (AURIN)
eResearch SA
Integrated Marine Observing System (IMOS)
Intersect Australia Ltd
Population Health Research Network (PHRN)
QFAB Bioinformatics (QFAB)
Queensland Cyber Infrastructure Foundation Ltd (QCIF)
South Australian Broadband Research & Education Network (SABRENet)
Tasmanian Partnership for Advanced Computing
Terrestrial Ecosystem Research Network (TERN)
V3 Alliance (V3)
VicNode

Victorian Life Sciences Computation Initiative (VLSCI)

Group C: Universities

Australian National University (ANU)
Curtin University
Flinders University
Monash University
RMIT University
Swinburne University
University of Adelaide
University of Melbourne
University of New South Wales
University of Queensland
University of Technology Sydney
University of Western Australia
University of Western Sydney
University of Wollongong
Victoria University

Other University-based Entities:

Council of Australian University Directors of Information Technology (CAUDIT)
Council of Australian University Librarians (CAUL)
Group of Eight (Go8)

Group D: Other (Funders, Industry, Individuals)

Australian Research Council (ARC)
National Health and Medical Research Council (NHMRC)
Fujitsu
Dr Markus Buchhorn
Mr Peter Nikoletatos
Professor Mark Ragan

B. Consultations

Consultations (face to face or telephone) were conducted with the following organisations or individuals. In some cases submissions from the same individual/organisation were also made.

ARC Centre of Excellence for Environmental Decisions, (Hugh Possingham); Australian Antarctic Division; Nathan Bindoff; Paul Bonnington; Ann Borda; Lindsay Botten; Sarah Brown; Rob Cook; Peter Elford; Ben Evans; Rhys Francis; Ian Gibson; Anne-Marie Lansdown; Stephen Manos; Clare McLaughlin; Glenn Moloney; Steve Quennette; Ron Sandland; Mike Sargent; Arun Sharma; Ian Smith; Liz Sonenberg; Neil Stringfellow; Peter Taylor; Alf Uhlherr; Ross Wilkinson; Alan Williams; Dave Williams; Bill Yeadon.

Also consulted through the Report preparation period were the Review Panel, and staff from the Australian Department of Education.

The above listing of individuals does not include those attending broader consultation meetings in Brisbane, Canberra, Sydney, and Melbourne.

APPENDIX THREE

Abridged Summary of Stakeholder Responses to the Terms of Reference

In the full Status Report provided by Professor Cochrane, a complete analysis is provided of all stakeholder and provider responses to questions against the Terms the Reference, along with a copy of those questions. This Appendix provides a summary of that analysis, which has been prepared by the Department. It follows the structure of the full Status Report, but combines analyses for eResearch Infrastructure providers with eResearch infrastructure stakeholders under each sub-heading.

I. Appropriateness

Market Failure

The majority of the stakeholder respondents consider that the Australian government investments have been a response to market failure.

For some, including the research agencies, most of the research centres and institutes responding, and the Academies (with the exception of the humanities and social sciences), this view is generally expressed. It is also held to be the case by most of the universities and individual respondents. The seven elements of the Capability are susceptible to some differentiation here. Some stakeholders provide a view categorised according to the elements of the Capability, while others comment on those with which they have had the greatest familiarity.

A general inference is that the peak facilities (NCI and Pawsey), most certainly would not have developed without national investment, and the services provided across the sector by ANDS, the AAF and the NRN connectivity, would similarly not have occurred. In the case of the development of tools through NecTAR and data storage and related services through RDSI, it is possible that some localised instances of these activities may have taken root, but in more localised and "subscale" forms: In the words of the Medical Research Institutes –

> due to decentralised development, it is unlikely that national standards of availability, management and sharing policies would have been adopted.

The stakeholder community was encouraged to express a view about the possibility that the relevant projects and services may have developed at the initiative of the relevant communities. The consideration of this question reaffirms the view that national initiatives are clearly necessary in the absence of evidence that the services would have developed in that more localised and decentralised way. In some cases the speculation is offered that some services would have been undertaken, but would not have reached the appropriate scale, been slow to develop, and/or been patchy. In other cases the view is more unequivocally negative, i.e. that there would have been little or no development. In fact, argues Geoscience Australia -

> the investment in eResearch infrastructure created a new market rather than responded to any market failure, to allow science innovation and evolution that previously could not be fully realised.

A more general point which emerges as stakeholders consider the reasons why the projects and services would not have developed in the relevant communities, is commentary on the commercial availability of capabilities for at least part of what has been provided by NCRIS. This theme also develops in consideration of some of the other questions, and it is important to note that this is a topic that has changed and matured over the period of the NCRIS investments (2006-2014). Indeed commodity offerings in both compute and data storage were one of the significant themes in the *eResearch Australasia 2014* conference. In the future, it is clear that this will be a major focus for consideration and planning, and some of the stakeholder views expressed for the purpose of this Report are worth deeper consideration.

A detailed response came from the Australian Academy of Science (AAS, through its National Committee for Data in Science, or NCDIS), which included an analysis listing six reasons why the infrastructure would not have been developed in the absence of support from the Australian government. These reasons reference a number of contextual issues, including the limitations on the ability of research communities to use the Australian public granting schemes (as administered in their current form by the ARC and the NHMRC), a topic arising with a number of parties during the course of this consultation. The six reasons are recited here in full -

> 1) A lack of funding /resources. Generally ARC and NHMRC funds can't be used to fund software development and researchers are reluctant to allocate any funding to support data curation/publishing;
>
> 2) Lack of motivation by communities to support/prioritize some services (e.g., data/metadata capture and sharing, online analytical/modelling services) or to collaborate on the development of shared infrastructure;
>
> 3) Conflicting agendas between research institutions (universities, CSIRO) , state agencies, state governments, national agencies, discipline-specific centres/networks;
>
> 4) Limited awareness of the benefits or necessity of core infrastructure (e.g., single sign-on, metadata standards, data identification and citation services) that are needed across the board.
>
> 5) Many of the services developed (e.g. ANDS and NeCTAR) were not available via the commercial sector – and if developed via commercial entities, they would have been much more expensive.
>
> 6) Currently data intensive science is prohibitively expensive on commercial clouds due to online costs of data access.

When prompted to consider the way that the services may have differed if developed within research communities alone, the most frequently cited such differences are (less or smaller) scale, cohesion, speed, and total economic return through co-investment.

Points of note included that of one of the research intensive universities, which offers a view of Australian government investment in terms of the big changes in research that have been occurring leading up to and during the funding period –

> There are fundamental changes in the way research is conducted, leveraging new resources such as massive data stores, enormous computational power and new ways of communicating. These changes are analogous to the invention of the telescope in astronomy and the microscope in biology. New infrastructure is expensive and requires a matching shift in research culture and methodology. In order for Australia to remain competitive it is essential that we engage. Unlike the telescope and the microscope, e-Research technologies apply to almost all research, and thus can be shared across the sector. Australian government investment allows this to be addressed at the highest level, rather than in a piecemeal manner. Economies of scale, and sharing of expertise, are essential and are best resources by top down funding. Without this, some research communities would fail to adapt their practices, and would continue to use ineffective and outmoded approaches. (UQ)

The service provider elements of the Capability offer a view that there was no likelihood of market solutions to carry out the various services and facility development with which they were charged. In a number of cases this is articulated as quite clear cut. For example:

- The ANDS view that there was no likelihood of a market solution to improving research data assets;
- In the case of RDSI, the absence of the possibility of the existing eResearch service providers having resources to become collaborative data storage providers;
- The categorical fact that significant extensions of AREN as achieved by the NRN would simply not have occurred; and
- The fact that individual investments by research communities were unlikely to have led to the scale of achievement in the case of the NeCTAR research cloud.

Changing and transforming research

The Capability stakeholders support the view that the effect of the investment has been to change and/or transform research. Indeed, this is one area where there is substantial consensus (over 80% of respondents views expressed being relatively unequivocal).

Some stakeholders nominate the discipline areas where the transformative impact has been at its greatest, e.g. astronomy, climate science, earth system science, oceanography, environmental and ecological sciences, bio informatics and genomics, geochemistry and geochronology (AAS). The CSIRO response considers that –

> Areas of distinctive transformative value-add delivered through programs already apparent include:
>
> - Improved tools for predicting and managing Australia's climate, weather, oceans, water resources, mineral resources, biodiversity, ecosystems, health diagnostics and emergency events.·
>
> - Increase in Australia's agricultural productivity and biosecurity, through collaboration platforms and genome projects for key crop, livestock and pest species.
>
> - Establishment of world leading astronomy capabilities in Australia, through new infrastructure, antenna design, data processing, survey enablement and the resulting discoveries.
>
> - Transforming fields of research from qualitative to quantitative, e.g. humanities and social sciences, geological analysis and digital collections.
>
> · Increasing community engagement in research, including enablement of citizen science.

CASE STUDY: TRANSFORMATIONAL IMPACT ON RESEARCH

The VicNode response on the integration of NCRIS eResearch capabilities with institutional and Synchrotron facilities in the Clayton precinct is telling and is quoted here in full.

> *The Characterisation (NeCTAR) Virtual Lab operates on R@CMon orchestrating data collected and `curated in "Store.*" (of which "Store.Sync" is the first example), making this data seamlessly accessible to MASSIVE for analysis - the specialist HPC facility for image processing. There are similar efficiencies gained by the 3D requirements of the CVL desktop delivered through the Research Cloud rather than the GPUs of the MASSIVE HPC facility. "Store.*" leverages the MyTardis data management tool, developed through NeCTAR and ANDS contributions over time, that permit concurrent facility and researcher view of data management (currently with 7 active facility instances across Australia, covering over 10,000 data collections). "Store.Sync" made the Australian Synchrotron a pioneer - becoming the first synchrotron to progress their contribution to research activity to include the bundling of data access with beam line time. "Store.*" represents a maturing and consolidation of many MyTardis instances and new instances emergent from RDSI allocations, into a single data service. Through predominantly VicNode RDSI allocations, "Store.*" will host data from the MX1, MX2and IMBL beam lines (2/3rds of the Synchrotron data generation), but also facilities such as MMI (a service that has delivered an imaging service to over $25m of research grants in the last 2 years), and FlowCore (over $50m). Through integrating NeCTAR, RDSI and ANDS the CVL and "Store.*" become mature services of value.*

The NCRIS eResearch Infrastructure providers all argue that the investment has changed/transformed the

CASE STUDY: TRANSFORMATIONAL IMPACT ON RESEARCH

The Imaging Centre of Excellence considers that the three facilities MASSIVE, (NCI linked), the NeCTAR funded Characterisation Virtual Laboratory, (CVL) and RDSI have been "absolutely essential" to the success of the Centre:

> *These capabilities have provided a comprehensive, flexible platform for a number of researchers working on different research problems but requiring similar computing infrastructure to achieve their research goals much more efficiently that would have otherwise been possible.*

conduct of research. The changes cited include:

- A transformation in astronomy which could not be imagined without the enabling development of the NRN;
- Research practice transformation with the move towards data citation, and
- Rich examples of transformative impacts through the integration of computational and data intensive services, including at NCI.

Pawsey offers the evidence that the use of the flagship facility (at Murdoch University), in the early stages led to outcomes involving over a hundred merit approved science projects per year using the facilities, generating more than 150 peer-reviewed publications and supporting research with an annual value of over $25 million. ANDS also makes a lengthy comment, cited in the full Status Report, commenting on the high level of material difference it has made in transforming research.

It should be noted however, that in general, the Providers have patchy information on the extent to which specific research outputs as a result of their "Projects" have been tracked. The 2014 NCRIS Survey shows some scatter among the Providers on this question with Pawsey, NCI and NeCTAR self-reporting full or partial monitoring of publication data relating to their "Projects".

Fostering collaboration

The NCRIS investments strongly feature collaboration in both word and intent and reassuringly, almost all stakeholder respondents state that greater research collaboration has been fostered through the investments.

This consultation has sought to understand the extent to which collaboration other than that directly among researchers or research teams may have been enabled. In some of the consultations this conclusion was volunteered. VicNode, for example, commented that the operators of the infrastructure found themselves working with teams all over Australia, with whom they would otherwise not have had contact, and therefore not had the ability to share expertise and insights in improving research infrastructure services.

Other forms of collaboration mentioned include the elements of the Capability themselves, as in the interplay between advanced computing, data capture, tools and cloud environments, which, given that it was an exact intent of the NCRIS investments is a confirming conclusion, though as this Report goes on to show, significantly tempered by challenges in coordination and understanding, and not always coherent timeframes for the realisation of the objectives of the various elements.

Mention has been made by several universities and university groups, about the impact investments have had in bringing researchers and support groups together, and indeed to support groups themselves collaborating in a way that did not exist before in the support of research. Typically this involves institutional IT services and libraries, but necessarily in many cases the support functions have also witnessed stronger partnership with the areas of institution with research management responsibility.

CASE STUDY: <u>FOSTERING COLLABORATION</u>

The processes undertaken by NeCTAR to establish this part of the Capability were reported as having encouraged collaboration from the very beginning. Thus the establishment of the Virtual Laboratory and eResearch tool projects have been activities comprising significant research collaborations in themselves.

- In the case of the Research Cloud an example "was the enabling of an IMOS international collaborative activity, the Ocean Data Interoperability Portal (ODIP)", while the All Sky Virtual Observatory supports "a collaborative arrangement with Intersect for delivering high-quality research infrastructure and tools".

Most of the providers describe significantly greater research collaboration through the use of the services. Thus:

- AAF mentions examples from the CSIRO (collaborative crystallisation), as well as other areas;
- RDSI points to its use cases to demonstrate the fostering of substantial research collaboration; and
- NCI argues significant convergence between the integration of supercomputing with the growth of data intensive services, generating research collaboration through the growing use of community codes.

In an example of the NCRIS approach itself fostering research collaboration, NeCTAR states -

> A common response from Virtual Laboratory stakeholders has been that only when coming together to develop collaborative responses to the NeCTAR RFP did they realise the significant value created through such integrative online platforms which combine the capabilities available from each of the disparate groups.

Indeed, the intent here has been to foster infrastructure collaboration between modelling/simulation on the one hand and observation/data on the other. This has facilitated greater collaboration within and between disciplines as in the examples of those using the VLs: All Sky Virtual Observatory (ASVO); Marine Virtual Laboratory (MarVL); Virtual Geophysics Laboratory (VGL); Biodiversity and Climate Change Virtual Laboratory (BCCVL) and Climate and Weather Science Laboratory (CWSvLab).

Meeting Expectations

Both stakeholder groups and the providers were asked to comment on the extent to which expectations about the investments have been met from their own perspectives, those of their stakeholders, and those of the funding parties. However, in many cases—and certainly among stakeholders as compared to providers—there is a lack of clarity about what those expectations precisely may have been.

In approximate terms:

- A little over half of the Capability stakeholders responding to this question state that their expectations have been met or exceeded;
- Another 40% consider that their expectations have been met by some but not all of the elements of the Capability, or respond that some of the desired developments are beginning to show benefit, but have been slower than expected; and
- A remaining proportion of approximately 10% express general disappointment.

The large research agencies, CSIRO and Geoscience Australia in particular, express a view that expectations have been a “good match” (or have indeed been exceeded, as expressed by GA). The Medical Research Institutes indicate satisfaction with AAF, NeCTAR and RDSI but are concerned about difficulty in accessing AAF services via an affiliated University, and the “general” satisfaction with RDSI is qualified by the statement that “access to resources is yet to be fully realised.” Some of the technical achievements have received praise, with AARNet stating that the NRN has exceeded its expectations, specifically with the DaSHNet project conducted for RDSI.

There is something of a scatter among the University sector, with the majority of Group of Eight and ATN institutions considering that expectations had been met, but some other universities, not so. Some components are singled out for high praise, contrasting in some cases with a critical view for the same component. NeCTAR is an example of this, where although disappointment brought about by initial slowness has been expressed, the same component has been seen by some stakeholders as having shown exemplary engagement.

Those recording disappointment in terms of the meeting of their expectations included the Australian Academy of the Humanities (AAH), expressing concern that more investment in humanities and social sciences (HASS), is needed, and stating the view that -

> the implementation of the [Understanding Cultures and Communities] Capability in the 2011 Roadmap remains an urgent need, including the cross-cutting Digitisation Capability. Humanities researchers need a national collaborative infrastructure that can connect dispersed social and cultural research data, make both old and new data discoverable and reusable, and allow analysis of large and diverse data sets.

In some cases the response about meeting expectations evokes a reflection on the implications of experience so far for future framing of investment approach and implementation methods. For example the ARC Centre of Excellence for Climate Systems Science, in the context of a concern that a major evolution in expectation has not been well managed, states a need for both a longer investment horizon and a better understanding by our research community of "the challenge we confront as a nation on management of data and the connected services associated with data, or the Capability associated with high performance computing. High performance computing is still central to several key grand challenge problems".

In the case of the response from IMOS, after stating that NCI and ANDS have met expectations, (in contrast to others), a caveat about ANDS is registered as follows -

> In our experience, ANDS is a funder of projects aimed at enhancing existing research infrastructure rather than a creator of infrastructure per se. The projects we have undertaken with ANDS have been successful, so the answer from that perspective is "yes". At a more fundamental level, the Department's approach of investing in eResearch 'infrastructure' projects that then run granting processes to allocate their funding should be scrutinised as part of this review process.

In terms of meeting the expectations of stakeholders, that is in this case the "stakeholders' stakeholders", the sense of the responses is that there is significant satisfaction with the meeting of expectations through the investments, but often accompanied by a concern about skills and expertise to exploit the new services and assets.

Further, as one individual states, the expectations of a key stakeholder—the Commonwealth—were clearly not met in the case of the decision to wind up ARCS, and the subsequent separation into the components NeCTAR and RDSI. To the extent that this process has had a legacy of some concern, confusion, and criticism of some specific subsequent service delivery delay, it is worth reflection.

The expectations of the providers themselves have been met. ANDS comments on its satisfaction with progress in achieving "the directions set out in 'Towards an Australian Data Commons', while the AAF remarks that it has become self-sustaining and more integrated following the initial modest investment. NCI states that it has achieved strongly in the context of its expectations, and NeCTAR qualifies its response in relation to some slowness in particular service development, partly due to technical complications. Thus its "unprecedented" infrastructure experienced some challenges in deployment.

The providers' views of the way their stakeholder expectations have been met vary as between each other, but are reasonably consistent with the views of those stakeholders about the providers. Both ANDS and AAF conduct satisfaction surveys providing feedback. The AAF reports average ratings of 3.75 to 4.3 on a scale of 1 to 5. The NRN considers that the expectations of a significant number of stakeholders have been exceeded. In some cases the projects have been able to "over deliver", examples being the Greater Sydney Basin Network and Greater Perth Fibre Ring.

NCI describes its stakeholders as comprising the Australian government, users of the supercomputer facility "in isolation", partner organisations and associated research communities, such as astronomy, geosciences, and industry. In all cases expectations have been met and there is further information provided about how

The providers also believe that the expectations of funding parties were met.

II. Contribution to Technology

The question about contribution to a platform of ICT support for the research sector has an inherent difficulty. This is contained in the way that the word "platform" is understood and interpreted. In considering information technology based services, this is not an uncommon ambiguity. The word "platform", can be interpreted by some to mean the one application, or presentation layer, by others to mean one "system", and by others to mean a commercial offering, a brand.

Nevertheless, most of the stakeholder and provider responses sought to address this question, and divided their responses as prompted into considering contribution to a platform "as a whole", and "in parts or selected disciplines".

As a Whole

Many of the stakeholders consider that there has indeed been a contribution to a platform as a whole. The CSIRO states simply that "The NCRIS eResearch services have delivered a reasonably complete platform of ICT support suitable for the whole research sector". Similarly, other research agencies, service providers, and universities express such a view, though Geoscience Australia notes the importance of understanding that "the platform continues to evolve and develop as there is not yet a defined 'endgame".

There is a strong sense that the notion of delivering a platform should be understood as looking at the synthesis between the eResearch Capability investments, together with the integration role played by others, particularly state-based eResearch service providers. The need for integration is a recurrent theme in this Status Report.

Again, some elements are differentiated in terms of the notion of contributing to a platform, although there is no consensus on this. The University of Western Sydney, for example, considers that it is hard to see a "Platform" as such, stating that for the majority of "long tail" researchers it is very difficult to get a whole picture of what is on offer.

All providers consider that such a contribution has been made. The AAF has become part of the research infrastructure landscape with 130 services connected, well beyond the University sector, and including facilities such as Garvan. The AREN is a fundamental component; and the NecTAR Research Cloud and National Server Program comprise a national computational platform "supporting highly diverse computational needs across the full breadth of Australian research"

Both NCI and Pawsey comment on provision of multidisciplinary support, with NCI reiterating the strength of its infrastructure integration -

> it is integration of the computational and data infrastructure, overlaid by expert services, that allows NCI to provide a highly effective, high-performance research-computing platform for the sector.

RDSI points to the storage of data from every Field of Research, and goes on -

> As was envisaged at the outset, the RDSI project has transformed the capability of the sector to store research data and make it available for collaborative use. Before the project, there were few, if any, easily accessible collections of data for collaborative use by the sector.

The response from ANDS takes a somewhat different tack, arguing that although it can be seen as an investment in ICT, it is equally importantly an investment in research data, and in doing so "has complemented significant state and institutional investment in research data infrastructure" enabling it to be more effective. Apart from the network infrastructure itself, it is difficult to think of a component of the Capability that has permeated so widely across the nation.

In Parts or Selected Disciplines

Some stakeholders have responded with views about relative differences in the uptake of, and therefore the value associated with, the contribution to a platform or platforms for research. Again this is

differentiated by component or project. As the Population Health Research Network (PHRN) points out "projects/services by their nature may have contributed to some disciplines more than others", a view echoed in several other responses.

Indeed the contribution from CSIRO ranks the discipline areas which have achieved the greatest value from the NCRIS eResearch investments, in the "approximate" order:

- Environmental and earth sciences;
- Physical sciences;
- Biological sciences;
- Chemical sciences;
- Engineering; and finally,
- Humanities and social sciences.

The CSIRO contribution goes on to say that opportunity for further benefit would appear to be strongest in "environmental, biological and humanities / social sciences, with variable uptake reflecting different levels of inherent IT skills within these disciplines".

There is also a detailed consideration by the AAS of the disciplines that have been served, viz -

> In some disciplines, the projects/services have contributed to a comprehensive ICT platform. For example, at NCI the combination of: petascale computing; fast access to spinning disk provided through RDSI; data discovery through ANDS; and support of virtual laboratories from NeCTAR (Climate and Weather, Virtual Geophysics Laboratory) have established the foundations for an integrated platform to perform data intensive science at scales never before possible in Australia, if not globally.
>
> However, the investment to date in national eResearch facilities/services has tended to favour certain disciplines (astronomical and space sciences, geosciences/geophysics, climate sciences, biological sciences (genomics), protein crystallography, ecological sciences, marine sciences). It has mostly supported those communities who are large, well-organized, have a common goal/problem, generate "big data" automatically via shared instrumentation (satellites, telescopes, synchrotron, sensor networks), use computational modelling as a research tool, and employ similar community-wide services/workflows for processing data.
>
> Research sectors in which there is significant potential for expanding the uptake and development of eResearch infrastructure include: humanities and social sciences; economics; engineering; built environment and design; materials science. These comprise communities with multiple, smaller, sub-disciplinary groups and sub-disciplinary problems and research methodologies. Generic approaches to ICT services are less likely to be successful and more tailored approaches are required. Some of the NeCTAR eResearch tools provide good, successful examples of such focussed communities, tools/methods and research data formats (e.g., OzTrack, FAIMS, Aust-ESE).

In some cases deficits rather than contributions are highlighted. Thus ANU suggests that social sciences need a better national technology platform, and several others comment on the lack of support for humanities, echoing the sentiments of the AAH.

Health and medical research receives some mention in this context. Intersect[9] considers that this group of disciplines could be better served, though some progress is noted by the medical research institutes themselves, principally through the activity of state-based eResearch providers on the east coast. Intersect states that in -

> NSW, Queensland and Victoria, Intersect, QCIF[10] and VicNode have been proactive in addressing this gap (including through a recently successful bid involving MRIs for a National Data Facility for Medical Research).

[9] Operator of the Sydney Node of RDSI and NeCTAR.

[10] Queensland Cyber Infrastructure Foundation, operator of the Brisbane and Townsville Nodes of RDSI and NeCTAR.

Complementarity and Co-Investment

In answering the question about the Australian government contribution intersecting with or complementing ICT support funded by state-based programs and/or institutions themselves, every responding stakeholder (on this question) identified that this has occurred.

This includes such complementarity and co-investment as occurred: within and across large agencies; with research institutes; with the eResearch service providers; through AARNet; through NCRIS domains; and through all universities. Some of the examples are well-known, and indeed co-investment was sought as a policy objective in the NCRIS strategy nationally.

Responding that such complementary investment has occurred, does not mean that this has been a straightforward process. The actual implementation of such co-investment arrangements has carried with it significant overheads and complexity.

All seven of the Capability elements describe funding complementarity. In the case of AAF, this has been a critical sustaining element allowing the initial investment to become a production (and self-sustaining), facility. For RDSI (via the RDSI Nodes), state and institutional funding are already combined in a leveraging process, and in the case of NCI we see research intensive universities complementing campus and regional computational support with peak access. NCI has worked with the ARC to provide the procurement of a service rather than infrastructure. Pawsey makes the comment that the research agencies (CSIRO, GA) and the Bureau of Meteorology (BoM) have been able to get strong complementarity in what they do, though NCI notes some challenges ("hard barriers") in seeking optimum leveraging effects with BoM.

One particular investment, ANDS, has seen widespread leveraging within institutions, quite often within service provider budgets (such as libraries). Co-investment data furnished by ANDS suggests that at least $3.7 million of institutional funds has been invested in programmes. Some research agencies and universities that have benefited from ANDS programs report co-investment ratios of 3:1, and significant "post investment", i.e. in data management following the implementation of projects initially co-funded with ANDS. CSIRO alone estimates this to be $3.5 million on major projects and it seems reasonable to extrapolate from data from some universities that the post investment activity stimulated by the ANDS process, could total over ten million.

Leveraging Plans

[*Note* that only Stakeholders were asked the question about plans to leverage the NCRIS eResearch infrastructure].

Almost all of the respondents to this question indicated that they would indeed, or are, seeking to leverage the national eResearch investments to date.

In the case of CSIRO, this is planned with particular reference to Pawsey, NCI, ANDS and NRN, with the agency seeking ongoing support mechanisms for selected RDSI funded collections and NeCTAR VLs. Research Institutes such as AIMS and ANSTO expect internal demand to grow and therefore the need to leverage to develop, with ANSTO answering in relation to specific components AAF, NCI, RDSI, NeCTAR. The medical research institutes also indicated that several institutes will undertake such leveraging.

As may be expected state-based eResearch service providers indicate that this has been happening and is planned in future. Thus in the case of QCIF, it is already happening; Intersect states that it has been happening and "we will continue to do so". However clarity about exactly what is planned is important in committing to co-invest; eRSA[11]: has such plans "definitely", while AARNet describes such leveraging, for

[11] eResearch SA, operator of the Adelaide Node of RDSI and NeCTAR.

"decades to come", through extending reach, additional services based on NCRIS achievements, collaboration with Big Science Projects, integration of AAF with cloud.

Similarly with universities: Melbourne indicating current practice in response to investments viz, the University -

> continues to leverage nationally-funded eResearch infrastructure by co-investing in and developing both physical and intellectual (people) infrastructure.

UWA indicates leveraging mainly with Pawsey, while UQ describes its colocation of facilities with NCRIS funded infrastructure at QCIF thus -

> UQ has already chosen to site new computational infrastructure with QRIScloud, namely a new data *intensive computer called FlashLite (funded by the ARC LIEF, QCIF, key universities and CSIRO). High speed connections between FLashLite and QRIScloud will support streaming large data collections into the machine for high performance analysis. It is likely that further investments in cloud infrastructure will leverage QRIScloud.*

Most of the other universities indicate similar intentions, with the University of Wollongong making the point that local and national investment are needed in partnership for eResearch.

However, the point is also made that plans for future leveraging are dependent upon clarity about future national funding with which to engage in such leveraging. Some describe the limitations and constraints on making leveraging investment decisions. This is particularly a concern in research institutions with constrained budgets.

III. Governance

Governance Arrangements

In general, this is an issue where there is considerable diversity and variation in the comments expressed, with many respondents drawing comparisons between various elements of the Capability. The views of the responding community range across the full spectrum from general satisfaction with the appropriateness of governance arrangements to significant concern. In many cases the responding entity describes an element of the Capability with which they are particularly familiar. The opportunity has been taken by many to raise the significant issue of national cohesion.

The majority of responses express general satisfaction with governance or satisfaction with a significant proportion of the elements of the Capability, accompanied by a concern about coordination, consultation and communication, particularly among the intersecting roles of the elements.

Across the range of views the most consistent theme emerging is the one stated succinctly by the Medical Research Institutes that the services should be "brought under the one umbrella". A similar view from the V3 Alliance proposes that a "single strategic governance body for research infrastructures, with a supporting secretariat, would be highly desirable", citing the example of the eInfrastructure Leadership Council in the UK. The need for greater coordination is echoed in the response by Geoscience Australia, concerned at the lack of a National HPC strategy.

As might be expected there is significant comment about the way that the governance model that has been generally pursued[12] has led to quite lengthy periods in which contractual and competitive processes have been seen to slow down the development and operation of some of the new facilities. One (ARC Centre of Excellence) respondent also suggests a cross-Capability governance, at the "level of AeRIC".

There are however variations on this theme. Acknowledging delays, one of the responding service providers considers that the process of extracting institutional agreement to action might be the main cause of delay. Thus according to VLSCI -

> the time it has taken to deploy some services might indicate governance issues, but these may be more related to the laborious process of obtaining institutional agreement that seems to result from the Australian research infrastructure funding model

A more top-down perspective would be that, where there was clear understanding and alignment of all the relevant parties in the Lead Agent arrangement, projects were likely to be well implemented. In cases where the role of Lead Agent has not perhaps being exercised with the intended level of nationally-taken perspective and responsibility, or where staff selection (or delays in recruitment) has also been an issue, the project concerned has been more likely to show evidence of problematic implementation. In one case, the AAF, the entity responsible for implementing one of the elements of the Capability changed three times before stability at a production level was achieved, and in another, a series of delays and evidence of some friction between the project's centre and its outlying operational areas which carried out project implementation (which became the structure of RDSI), affected delivery time very significantly. A negative view came from more than one state-based eResearch service provider, about the level of difficulty experienced in the relationship between the Nodes they operate and the project centre in the case of RDSI, as well as frustration at seeking to influence this.[13]

Many of the university inputs indicate a view that governance has in general being adequate and appropriate. Not all however: one university considers the AAF and ANDS projects to have been well governed, but others relatively poor. Concern about the accountability of project leaders is referenced in this response.

[12] ie, selecting and negotiating with the Lead Agent, establishing a board, appointing staff, and resolving contractual issues at the level of both government-to-Lead Agent, and within the project.

[13] It should be noted that the Super Science RDSI project was the subject of separate review in 2014, which included governance and other issues.

The response from the Australian Academy of Science raises the question of Board and Committee representation, considering that they should have greater diversity. In a similar vein, one service provider nominates the need for greater accountability and feedback arrangements with end users, and operator presence on boards. A separate observation describes the problem of uneven capacity or talent in the process of leading projects. In some cases dependence on the individual preferences and capabilities of a project director have made the project vulnerable

All seven elements of the Capability consider that the governance arrangements have been appropriate, although the commentary is somewhat nuanced.

- NCI considers that its governance arrangements are "entirely" appropriate, citing investment by partners as ensuring commitment;
- RDSI nominates two caveats, concerning relations with Government on the proposed process for administering ReDS funding, and more recent communication processes.
- NeCTAR states that its governance arrangements need to be revisited as it transitions from development to operational support, including "the right balance between co-investing representative governance, expertise-based governance and sector-representative governance".
- Both ANDS and NRN show evidence of having thought significantly about governance to ensure success, and believe that this has occurred.
- NRN deems that its governance arrangements have worked extremely well. It was multilayered in that there was an overarching steering committee for the project, as well as individual steering committees for each subproject.
- The AAF comments that it is governed well by its members and must also meet compliance requirements as an incorporated association.
- Pawsey refers to appropriate construction and implementation phases of its project, and makes reference to the need for the iVEC Board, charged with the responsibility for governance of Pawsey as a piece of national infrastructure, to balance partner interests with the discharge of that national obligation.

IV. Take–Up

In the views of many large institutions and networks, take-up has been wide and significant.

This tends to have been stronger in some disciplines, naturally, and in situations where some of the funded projects and services have matured only recently it is thought difficult to assess a complete picture of uptake. This comment, for example, from the AAH -

> It is early days for the HASS-focussed NeCTAR funded projects and services – it is difficult to gauge at this stage what the uptake of the services by researchers will be and the extent to which the investment in these projects changed or transformed the conduct of research.

Larger institutions tend to depict generally positive results, this being a view shared by most of the research intensive universities and CSIRO and Geoscience Australia. The University of Melbourne notes that in the context of generally good take-up, a better balance of academic and "operational/technical" input would have improved governance. It also notes that the

> structure of recent investments has meant that the burden of navigating the informal connections between compute, cloud, and storage generally falls on individual researchers, rather than being systematically addressed by the design of provider offerings.

As with much of the input to this Report a lot of the stakeholder commentary differentiates elements of the Capability. One particular comment offered on the services from the AAF, is a commendation indeed - "a service that now appears to provide the capability that everyone likes in an IT project – it just works."

The views of the state based research service providers tend to greater variability. While the Tasmanian provider comments on the "phenomenal" uptake of NeCTAR cloud services, their equivalents in Queensland, New South Wales and South Australia comment on a more variable picture. Reference is made to the major cultural change in the adoption of some eResearch practices and infrastructure, and therefore its inherent slowness, although a decisive impact has been felt from the start that has been made in Australia through these investments.

Another issue for those working in medical and health areas has been the dependence of the MRI sector on university partners for open and direct access to eResearch facilities. This has not always worked well. Indeed, greater seamlessness is a major theme across the sector across a variety of disciplines and science.

Other areas for improvement include significant enhancement of communication, outreach, and engagement by service providers, reference to the significant need for local levels of technical skills and skills development, the fact of delay in itself in reducing take-up, (RDSI, Pawsey, some aspects of NeCTAR). Referencing the need for greater additional local skills' availability also implies significant additional investment, particularly because, in the words of one university "there is an expertise gap between current research practice and eResearch".

One research agency in its response on this matter raises a significant concern about growing dependency for its operational activities on the achievements developed through the investments in recent years -

> A significant issue for GA is that the success of the eResearch services has resulted in the transformation of a number of GA science activities to the point where the NCI and RDSI is a dependency.

Another issue inhibiting take-up relating to the characteristics of the investment funding flows, is nominated by the Australian Institute of Marine Science, and others, namely that uncertainty about future funding itself has a dampening effect on the extent to which research managers are willing to develop a new approach with no certainty of continued revenue support.

All seven components of the Capability consider that their infrastructure has been well taken up:

- The response from ANDS describes three forms of infrastructure, being national research data infrastructure, institutional research data infrastructure, and collections. ANDS has also run services supporting usage.
- The peak facility providers, NCI and Pawsey both comment on oversubscription demonstrating unmet demand.
- The AAF describes comprehensive take-up by the universities, CSIRO and state consortia. It also describes good integration with RDSI and NeCTAR in enabling Enhanced Client or Proxy access in some cases to non web services.
- NeCTAR points to strong continuing growth in the take-up of the Research Cloud, both directly and indirectly. There is evidence that a broad spectrum of disciplines is being served, and the NeCTAR submission provides illustrative and graphic material as well as a reference to collaboration with Microsoft, Amazon, and the AAF to investigate AAF-based access to the Azure Cloud through NeCTAR interfaces.
- RDSI comments that at the time of its submission, 11 petabytes (PB) of storage have been commissioned and expresses confidence of achieving the project goal of 75 PB.
- The NRN states succinctly that "by its nature the infrastructure developed by the NRN project has been taken up as soon as it was delivered. There are no barriers which inhibit take up".

Barriers Inhibiting Uptake

The main commentary provided here can be categorised in three ways:
- Barriers resulting from deficits in skills and expertise across the board;
- Those resulting from the fragmented nature of the Capability components, which have both structural and communication-and-engagement elements; and,
- Those arising from an uncertainty about the future and sustainability of the funding approach.

A related issue is that of branding and ready identification of the investments. Another of the ARC Centres of Excellence suggests that it has been hard to recognise investments, so that clearly branding has been an issue (CEED), and lack of awareness is also described in some of the University responses, reflecting that communication and engagement could have been executed better.

The notion that the conditions of funding have inhibited the translation of capital into ongoing sustainable approaches to service, and uncertainty about sustainability more generally, is widespread. Of course it is understood that significant investment could only be made available under these constrained conditions, but these nevertheless later translate into a sub optimal uptake in the view of some responding institutions and research domains. This is expressed by one NCRIS domain (TERN) when it states that a lack of knowledge about the future is itself a barrier -

> as Institutes and research groups don't want to invest considerable amount to time and effort to know and start using the eResearch capabilities when they are not sure how long the entire program will last.

It is important here to convey the sense that the responsibility for addressing these barriers is a shared one. While responding institutions and service providers may well observe that stronger national coordination and a more sustained approach to funding would have helped reduce barriers, it is also clear that research leadership (however that is exercised across the sector, in the context of the institutions in which it is required) must take responsibility for underpinning and supporting changes in research at the institutional level. There are significant differences among Australian universities in the way they have tackled eResearch infrastructure, and exploited the opportunities offered by the Australian government's program of investments, and by 2014 the contrast among some institutions as a result of those different approaches is evident.

The providers have given varying answers on the potential for improvement in take-up. NCI considers that it would be difficult to improve on the quantitative level of achievement to date, however, the NCI response does nominates two significant barriers to take up, being lack of staff in user support and HPC

innovation as the first one, and as the second, a criticism of the way that resources have been provided with an emphasis on “data in the abstract”.

The response from ANDS implies the answer that it would be hard to improve on the achievement to date. On the issue of barriers, NeCTAR acknowledges the initial adoption hurdles, and the long development phases which slowed the early uptake by users, and states that in its case the Project necessarily took time to ramp up, but will show significant acceleration.

RDSI expresses the view that operational funding might be invested in the acquisition of public cloud archiving which would free up capacity for more research data as and when required. The AAF advises of significant possibilities for improvement from its present operating capability. The AAF and RDSI have both, for different reasons, been affected by the technological complexity of what they were attempting. The AAF considers that it could have solved some of the problems with further funding. In its response RDSI lists seven factors thus -

> barriers have included the novelty of the project, the unfolding complexity of the problem, the restrictive nature of EIF funds, the relative immaturity of nodes as service providers, some difficulty in engaging other service providers, the difficulty in establishing communal decision making amongst independent legal entities, and the technical difficulty of the project.

v. Lessons Learned

Respondents were asked their views on lessons learned about the optimum delivery of this kind of infrastructure, as well as the extent to which any of the issues in delivering infrastructure are systemic in nature or otherwise, and there is a rich set of responses. It should be noted however, that there seems to have been an uneven perception of the meaning of "systemic".

Some of the responses interpret the word "systemic" as applying to the system of investment and the NCRIS approach. Some organisations took the opportunity to offer "big picture" views of Australia's strategic direction, and the way it chose to develop and specify that direction. Others had more specific commentary and advice on improvements that might be made in the actual implementation of an investment approach.

Higher Level Commentary

Several research intensive organisations offer an overall view about what is being achieved. The Group of Eight comments that "NCRIS facilities are seminal contributions to national research infrastructure. By their nature they are not small or temporary, but are significant and enduring and support research into the future. This requires an ongoing investment strategy.

One of the State providers suggests that operational maturity has developed from the approach, as it has built a collaborative method which reflects the way that research needs to be supported. They also argue that the programs have begun to develop the workforce necessary to bridge the gap between technology and research practice.

There is reflection in some of the responses about the approach in terms of the funding allocations there is also in some commentary an overall acceptance of the idea two categories of need. The ARC Centre of Excellence (ARCCSS) describes these two as "peak" and "majority", which must be met in different ways.

There are some contributions here about the path that Australia took in developing the early engagement to construct the approach. Thus for example, on one view (AARNet), the broad-based initial engagement was a good start, but some mechanism of continuing that level of engagement should have been considered. A parallel opinion is expressed by PHRN, put simply, that a collaborative national process "delivers a better eResearch infrastructure outcome".

One perspective from the experience of AeRIC, is that for the funds providers, it would, given the Lead Agent model, have been preferable to have greater certainty of buying in to the idea and actuality of a collaborative infrastructure as needed from the partners receiving the investment. Conversely from the perspective of the research sector, more transparency about how budgets are developed within government, and how some of the initiating decisions are arrived at, would be helpful.

Advice on Improvements

Many of the lessons learned are about how the process and the approach might have been better. Included here is the idea of better engagement (including upwards) in policy-making, particularly to influential decision-makers who are outside the direct line of ownership of the infrastructure, as well as outward to the research community; greater attention to communication among the silos, and particularly thinking from the outset about interoperability between services such as the ANDS meta data stores and RDSI storage, or between AARNet CloudStor and NeCTAR Research Cloud nodes. There was also a view that some of the eResearch entities should not have functioned like "granting agencies". One expression of this was in the response from Astronomy Australia Limited (AAL) -

> Some of the eResearch entities function more like granting agencies (rather than providing core eResearch infrastructure or expertise) and use a range of different mechanisms for awarding grants/resources. This model does not appear to be aligned with the top-down strategic approach of the NCRIS program. Indeed, the eResearch capabilities appear to have been conceptualised more from the bottom (i.e., infrastructure layer) up, rather from the top (i.e., research outcomes/objectives) down. As a consequence, the engagement of research communities, such as

> astronomy, in prioritising investments is retrospective and reactive, rather than proactive from the start.

Similar sentiments have been expressed by another NCRIS domain, IMOS. And ANU offers the additional view that the grant awarding programs run through ANDS, RDSI and NeCTAR, involved a disconnect "between governance and user communities". This then led to a greater level of difficulty than necessary in creating integrated services for some of the service providers seeking to knit together the combined offerings of these components of the Capability. This view is echoed by other research institutions. Comment is also made on the opportunity cost for cross project collaboration being lost due to projects maturing separately. This might have been improved by an overarching governance approach, and closer alignment with ARC and NHMRC grant programs (University of Melbourne).

Another theme in the responses on lessons learned is to consider, in the later part of the investment period, (i.e. the last two years), the emerging maturity of commercial alternatives. The view is posed (UWA), that had the investment period been in place now such commercial alternatives would be considered. Conversely, in the VicNode contribution, the statement is made that the $/core and $/TB of compute and storage are cheaper than commercial cloud offerings of equivalent products.

From the provider point of view, the lessons learned tend to be expressed in variable ways according to component. A common thread for some of them is the question of better alignment with each other. NeCTAR offers the observation that distributed implementation has been an issue, with its submission noting "while the sector valued the openness and transparency of the NeCTAR RFP processes, we would recommend that a future VL–like program include a strong sector-engaged facilitation process to support the development of community proposals, especially in partnership with the research domain oriented NCRIS capabilities".

NeCTAR also underlines comments made by other stakeholders about the implicit tension between RFP processes and partnership development. There are essentially two takes on this issue of collaboration in a competitive environment.

- The first is the macro issue of developing shared infrastructure in a "system", that is essentially competitive and constructed to be so in terms of research funding and granting processes.
- The second, at the level of the development of the collaborative infrastructure itself, is a critique of the use of competitive processes in constructing what is to be a collaborative environment, and this is certainly one lesson worth reflection in considering future approaches.

The AAF considers that there has been a need for better leadership, project management, coordination and integrated vision across the elements of the Capability: "projects and their governing bodies need to be constructed such that cross project integration is a goal/performance measure that their success can be measured upon". The AAF also comments about coordination, and systemic problems in authorisation for the different NCRIS projects.

The NRN reports only positive outcomes in terms of the lessons learned, and lists them as follows-

> The importance of capable and committed leadership from the Lead Agent
>
> The important of an appropriately constructed project steering committee working closely and collaboratively with the Lead Agent.
>
> Mutually reinforcing governance arrangements which place responsibility for outcomes at the appropriate level.
>
> Clear and robust (fixed price, milestone driven payments, penalties for non-delivery) contracts between the Lead Agent and sub contractors.
>
> Clear and robust project management procedures including for risk management, change management and reporting
>
> Committed and engaged subcontractors with an ongoing stake in the operation of the built assets.

ANDS considers one main learning to be that ANDS as an infrastructure and data as a resource need to be thought of differently, in terms of the way resources allocation is conceived. Thus, the initial conception of "Platforms for Collaboration" imagined an ICT environment that supported other capabilities, but ANDS itself has become infrastructure which is in turn an "essential dimension to planning infrastructure investments. Data investments are not purely technical, and the value of data investments increase as usage rises, and are quite different to allocation of scarce resources".

Pawsey reflects that a significant learning has been about the poor fit between planning time frames for implementing a built facility, as well as the actual planning of compute and storage and the subsequent allocation processes. A specific issue is the challenge of dramatically different time frames for building as a project activity versus technology changes in ICT. Of particular concern is that the "Pawsey centre may well be unable to accommodate the SKA requirements,… not due to any deficiency in the skills of the people involved in the design of the building, but… simply a consequence of attempting to plan requirements almost a decade in advance of their expected delivery".

When contextualising their remarks about "systemic" challenges, provider comments reflect a view of the system (of research institutional organisations), and also the system of actually conceiving and developing the NCRIS program. An example of the former commentary from RDSI –

> Yes. The lack of agility by individual universities has meant that those nodes who have depended on universities for procurement, installation and operation have been significantly slower than nodes who are incorporated, and own and operate their own infrastructure

The most significant comment on the actual process of the program comes from NCI, and is quoted here in full.

> Yes. NCI believes that the delivery of e-infrastructure services has been adversely influenced by its conceptualisation almost exclusively from an infrastructure perspective, rather than from the perspective of advancing research ambition and outcomes. The implementation of some eResearch projects exacerbates this design inadequacy through governance that is insufficiently connected to the outcomes, an unwillingness or inability to prioritise investments for those projects that disburse monies to the sector, and an absence of an overarching vision that aligns with the advancement of national research priorities and NCRIS tenets of national, strategic, collaborative and world-class.
>
> It is suggested that this reflects the limited advice that has been available to the Department previously, and the need for a national e-infrastructure strategy, the architecture of which reflects the requirements of the entire research and innovation sector, given its increasingly central role.
>
> The relative lack of investment by the university sector, with notable exceptions, in high-end computing is problematic. The continuation of merit-based access being subsidised by the subscriptions of paying partners is contentious, particularly so in tough economic circumstances. There is a range of possible solutions, some of which require the research councils being at the infrastructure-planning table.
>
> It is arguable also that there exist problems in delineating between requirements that are national responsibilities and those that are institutional. This is evident in both the NeCTAR and RDSI implementations that have led to ongoing dependencies by being implemented in 8 nodes that span every state and territory (with the exception of the Northern Territory). Unless there is clear delineation between national and institutional responsibilities, and clear statement of purpose for the investments, the inefficient/ineffective use of monies at the commodity layer has the potential to weaken Australia's high-end capabilities that are so critical to its future competitiveness.

Recurring Themes

Many of the comments made in the responses on "lessons" are reiterations of points made throughout the contributions. Thus, concerns about:

- Communication, outreach and marketing;
- Integration and greater national coordination;
- Longevity of funding;

- The structure of some of the project delivery (e.g. the idea that RDSI might have been "centralised", rather than being based on state-based delivery);
- The separation of RDSI and NeCTAR;
- Slowness in some facility development; and
- Time pressures in the acquitting of funding, described in one comment as "unhelpful rushing".

The Department and AeRIC were aware that administrative overheads involved with the establishment of the arrangements, which were anticipated, turned out to become greater where the arrangements became contested, or choices made in the implementation (including staff recruitment) turned out to be problematic.

There is an obvious line of argument about the extent to which the separation of infrastructure from the research that it is supporting is itself a systemic issue. Clearly there are choices to be made here, but the problem of building critical mass through more efficient marshalling of funds for explicitly identified infrastructure (whether "e-infrastructure" or not), is a policy issue for every advanced economy. The most helpful approach may be to simply understand that this structural segmentation comes with the territory, and requires the building of processes and programs that minimise its downside.

One of the contributors, the V3 Alliance, provides a long list of 11 lessons, being: a need for clear frameworks; alignment of "funding parameters" with phased stages of an overall plan; coordinated roll-out; differences between projects and services being understood, (in terms of skills, governance and a range of matters); and addressing various deficits in the way developments roll out, listed in the full Status Report. Also mentioned in this submission are service delivery frameworks and associated funding needed after roll out; more structured learning from the past and overseas; greater attention to outreach; investment in training; invest in small "rapid projects" as well as large infrastructure; recognition of the "sustained requirement for software development and maintenance".

Closer attention to project management as a discipline, including the specialised branch (ie of this as a discipline) of this involving large project management and procurement is a clear lesson from the program of rollouts over the nine year period. There are also comments from stakeholders about the systemic nature of some delivery issues focus on alignment of activities (i.e. in the way that research and its funding is structured). Some also consider that the challenge in meeting the gap for the appropriate skills may be regarded as systemic.

There is some commentary about the way that separate contracting arrangements appear to be a systemic issue. Thus one NCRIS Domain (TERN) states that "some of the issues may be systemic, partly because each project has their own milestones and key performance indicators", while TPAC offers that "there are significant systemic issues in [the] e-Research Community, that are pervasive and to some extent demanded by the contractual arrangements".

While the issue of recruitment and retention of key enabling staff has been mentioned many times, (for example nominated as "the greatest challenge" by the University of Melbourne), a countervailing and more upbeat view has also been expressed. For example QUT considers that the growing evidence of mutual collaborative investment by universities, in research facilities in their regions, (Brisbane based examples being TRI and QFAB), has complemented the NCRIS processes, leading to new patterns and habits of research. If choices are made and resources available, research institutions can choose to invest in experts and developers that work across multiple projects, and in sufficient quantity to help meet the expertise challenge. This approach is mentioned in the Monash response.

One of the more comprehensive responses about "people", was from CSIRO, Thus, the comment runs…

> the main limitation in developing and operating eResearch services is the people. Delivering eResearch services requires specialised knowledge and skills beyond what is customarily learned in educational institutions or other employment sectors. Career path opportunities fluctuate in scale and lack clear definition, resulting in people moving into this field more by accident than design. Short term investments often necessitate short term appointments, which are not necessarily

> attractive relative to alternative employment opportunities. As a result there have been instances where institutions have allocated staff to NCRIS projects based more on expediency than aptitude. Enhanced skills, training and career track is a systemic issue. A related limitation was that the eResearch Capability became highly dependent on individual senior appointments, and on the roles played by a small number of key individuals driving the overall program from a truly coordinated, national perspective.

Some of the commentary on skills is more specific. An almost identical comment on a systemic tendency for research projects to hire researchers "to develop IT services instead of people experienced in providing IT services", is made by both Curtin University and CAUDIT.

Given the Australian approach over almost a decade, it is not surprising that those close to the implementation of infrastructure have significant and generally constructive commentary about how its implementation may have been better. A lengthy comment from the Australian Academy of Science is worth quoting in full, covering as it does the Australian administrative context specifically.

> There needs to be better overall synchronization and coordination of the eResearch facilities to: reduce duplication; improve integration and interoperability between facilities; reduce apparent competitiveness between facilities; improve outreach to research communities via a common marketing/information strategy across all facilities - so it is clear who is responsible for which particular services and who researchers should contact with regard to specific services.
>
> One suggestion is to re-establish an entity similar to the Australian eResearch Infrastructure Council which was disbanded two years ago. AeRIC was active in seeking to ensure integration across all of the individual programs and to minimize duplication. AeRIC also helped raise awareness of the activities between the individual programs/services.
>
> In the past, DIISR also facilitated meetings and cross-collaboration between the various capabilities (e.g., AuScope Grid, EMII, TERN) – ensuring harmonisation and sharing of developments. DIISR's past involvement in the eResearch Australasia Conference also facilitated outreach/awareness-raising and coordination of the various NCRIS/SuperScience activities.

VI. Comparisons, including Comparable Developments

Respondents were asked, considering the approach taken by Australia in undertaking these investments, how they would rate that approach by comparison with both international developments in comparable contexts and infrastructure development in other "industries" or sectors.

Comparisons

Australia's achievements to date are commended by the majority of responding stakeholders. Indeed some of the achievements are world-class, as in the example from Geoscience Australia concerning the development Data Cube providing coverage of the Australian land mass.

Stakeholders consider that AAF and ANDS are known leaders, and a particular point is made by one state provider about collaborative achievement under NeCTAR thus -

> Programs such as NeCTAR have achieved what has never been done internationally before - building a single infrastructure service that operates across multiple legal and state boundaries, to form a cohesive service offering, allowing researchers to easily collaborate across these boundaries (VicNode)

It should be noted that the providers offer comment only on the issue of international developments in comparable contexts, with no comment being made about other industries or sectors, apart from a comment by NeCTAR on its own interaction within the OpenStack "industry".

International developments in comparable contexts

Moving to international comparions, some offer a rather more mixed verdict. In some instances the view is expressed that Australia compares well, and in some cases some elements of the Capability are recognised as internationally leading, but this is qualified by the observation that this position can shift rapidly, highlighting the need for "continued investments to ensure a predictable, stable and continuous set of eResearch services on which researchers and other infrastructure providers can rely".

Some of the concerns about international comparison described divergences in meta data standards with choices made elsewhere. Thus ANSTO expresses the view that "a national eResearch organisation coordinating the development of standards and technologies following international initiatives could significantly increase and accelerate… benefits" It is unclear whether this comment acknowledges some of the integrating work being done in research data collaboration internationally.

Australia would be in a stronger position, capitalising on advancements made over the last nine years, if there were a solution to the issue of "jerky funding" (AAS). This issue is also expressed in terms of needing a better long-term vision nationally. There is some concern that these deficits may lead to Australia "falling behind".

Particular concern is expressed by the Australian Academy of Humanities that the nation is behind international benchmarks in HASS-

> The failure to invest here has kept the Australian humanities community out of collaborative international engagements, and their leveraging opportunities. In Europe, major HASS infrastructures include DARIAH (Digital Research Infrastructure for the Arts and Humanities), CLARIN (Common Languages Resources and Technology Infrastructure), CESSDA (Consortium of Social Sciences Data Archives), the European Social Survey (ESS), and SHARE (Survey of Health, Ageing and Retirement in Europe).

In terms of international comparisons, the commentary divides as between the peak facility providers and the other elements. Thus ANDS offers the comment that it is world leading, –

> When international developments in research data are being considered, Australia and ANDS are regularly cited as being leading approaches worldwide, if not best practice. The approach of having research institutions as central to research data is both unusual and highly regarded.

NeCTAR offers a view about pioneering work, especially with the Research Cloud –

> the only cloud-computing based research infrastructure operating at a national scale; and the only example in the world of a single OpenStack cloud fabric operating across distributed data centres operated by different organisations; and Internationally recognised within the OpenStack industry as a leading deployment of OpenStack

NeCTAR has been invited to join Europe's EGI federated cloud. The VLs have similarities with Europe's Virtual Research Environments, VRE program and the US Science Gateways program. There are potential links with European VRE developments for Horizon 2020.

The AAF states that it is in the top 10 as measured by connected services and identity providers internationally; with no one else having the complete University sector connected, and with a strong research orientation developed in a relatively sustainable and agile way. The NRN believes that it compares favourably with other national research and education networks, and RDSI is unaware of comparisons with its precise role.

Conversely, NCI considers that the approach taken has not given sufficient emphasis to HPC and that there has been an over emphasis on "data in the abstract". Australia lags behind in supporting higher-level functions such as software development and code optimisation. Similarly, Pawsey offers a lengthy treatment of the differences between the Australian approach to date, and elsewhere, particularly Europe.

Infrastructure development in other "industries"/sectors

There have been substantial issues of interpretation about this question. It was designed to elicit a view, from folk working within a research context, about their knowledge of the development of collaborative infrastructure in other industries. In some cases this has been the interpretation, but others interpret the question to mean the development of HPC Capability in other industries; or the development of commodity provisioning to the research sector; or the inference that what is sought here is a discussion on how the approach might have been different in developing a national program.

One respondent took the opportunity of the question to suggest that more rigorous international benchmarking of the Australian approach to research infrastructure might be undertaken, – exposing the Australian approach to more international view and review, in an ongoing way. It is clear that many respondents felt unable to answer, sharing the view of TERN of being unaware of comparable development in any other sectors.

However two organisations tackle the issue from somewhat different standpoints. Thus Geoscience Australia remarks that –

> Although HPC is being adopted by some oil and gas companies, few industries have the resources required to support such infrastructures. These industries are focussed primarily handling larger volumes of data and evolving traditional research approaches rather than the transformation of science that has been occurring through the collaborative research platform supported by NCRIS.

while the Medical Research Institutes make what seems to be a more general observation about the greater ease of working with industry provided services, viz -

> Overall, the view of MRIs is that industry has a more coherent, integrated and user-friendly approach than government-funded eResearch services. Government eResearch services are not seen to be sufficiently open and integrated, and the user experience seems secondary in government offerings. Marketing and implementation timeframes are also viewed to be suboptimal.

One matter raised by several is a wariness about the meaning of "Big Data", in commercial and research contexts respectively. The scale is completely different, although this is a dynamically moving area, and one should be wary of durable conclusions. What seems likely though as the world moves to Exascale compute and data operations, is that commercial dedicated capability will necessarily lag, based on its own efficiency and viability concerns.

VII. Influences and Constraints

Influences and Constraints (Stakeholders)

Budgetary and funding issues

As expected, a repeatedly expressed concern is about the periodic nature of the funding, combined with deadlines on expenditure requirements – the meeting of which was sometimes contradicted by delays imposed through the time taken to reach contractual agreement for some implementations.

There is the additional view that another constraint has comprised the limitations on expenditure for operational and support staff. Where this does not occur, where a project is unambiguously about the implementation of change through capital expenditure, achievement can be more straightforward. AARNet singles out the NRN in this regard, because of its simple funding arrangements and the way that the funds could be applied.

There is concern not just about the periodic nature of funding, but also about the uncertainty generated by knowing that future arrangements are not clear, inhibiting institutional responsiveness. (Of course the converse argument applies, namely, that leadership particularly in institutions with significant funding is, in some cases, lacking).

For operators seeking to meld various aspects of the Capability together to provide more seamless service for the institutions they are serving, some of the budgetary constraints involved with periodic funding, together with unexpected delays, have combined to pose very large challenges. This is further complicated where different rules for the application of different funding sources by the one entity introduce complexity and inefficiency.

There is also comment, not on a constraint, but rather an issue of cost effectiveness. Thus TPAC notes that the services can be seen to be cost effective when compared to the hidden overheads of researchers managing their own data collections, for example.

The response from ANSTO sees the principal constraints as a combination of budgetary, structural and technical, thus -

> A combination of all of the above [ie budgetary, structural and technological]:
>
> i While there is huge potential and benefits for eResearch projects, they are usually short-lived, meaning there is no ongoing funding for operational aspects. Also, due to the limited funding and short time-scales, eResearch projects are often delivered in prototype quality.
>
> ii Different disciplines have different approaches to eResearch. While there is a benefit for individual solutions, there could be a national program aligning and underpinning these activities.
>
> iii The lack of standards, a common technology stack and well-defined interfaces between eResearch service providers limits the impact and success of institutional projects to some extent. Also, there seems to be a gap between the national eResearch service providers (AAF, ANDS, NeCTAR, RDSI).

Inputs from the Capability service providers to some extent mirror those of the stakeholders. Generally speaking, there is little concern about the overall size of budgets. ANDS comments that while a larger budget would inevitably help, what is also needed is sector "readiness" –

> The budget was large enough to cause a national change in approach to research data, with an implemented research data commons, but length and reliability of investment is more important than size of investment, as institutions need to commit to long lived data and national services and support for such data should be appropriately long lived.

Mostly, provider budgetary concerns are about approach and balance. In terms of approach, RDSI, and NeCTAR comment on the constraints of EIF funding. This contributed to a slow ramp up in the case of RDSI, and in the case of NecTAR, despite the fact that overall it was well resourced, and attracted high co-investment, the EIF approach inhibited coordination of operational investment across programs, and

common services, and the approach led to dubious value proposition for a service provider to fund NSP nodes for a "national" research community, and finally inhibited Directorate size and efficiency.

AAF is the one Capability element that proposes that further investment is needed to develop to meet requirements now, particularly in more fine-grained control of authorisation; access to non-web services; connection to other authentication protocols and to other non-AAF users.

Structural Issues

There is a wide array of views stimulated by the question about structural issues.

There is some concern about the structural implications of seeking to leverage co-investment, particularly through state governments. Although this has its origin in budgetary strategy, it is argued that it leads to a change in goals, so that the Federal/State combination does not always align with the objectives of research as seen by (the Federal) government or institution as well as it might.

The issue of timeliness also crops up here. VLSCI makes the point that in this sector a delay of "four years in rolling out a service likely means that what is eventually rolled out is already obsolete!"

Comments are made on the structural approach to particular parts of the Capability, as well as subject areas that may not have had the focus that they would warrant as a research priority. On the latter point there is a view that health and medical research may have been, at least initially not as well addressed as some other disciplines. In the case of the comment on structure, CAUDIT notes-

> The decision to construct RDSI nodes across the country rather than one datacentre used by all was an unnecessary complexity that jeopardised the whole project and has restricted its capacity to deliver the desired benefits

Issues of discipline orientation are also raised in the response from the University of Melbourne, particularly relating to the level of organisation and readiness within those disciplines-

> Another symptom of the lack of a sustained funding program is that the investment in planning is also stop-start. Some discipline communities are mature enough to maintain a flow of analysis of emerging opportunities in the form of discipline roadmaps. Many are not, for a variety of reasons. This is an area where modest national investment, to sponsor continuity of both "vertical" (i.e. discipline focussed) and "horizontal" (i.e. enabling technology focussed) standing committees would ensure greater responsiveness, and more balanced addressing of emerging research community requirements, as major ongoing schemes become refunded.

Some respondents regard the issue of skills and training as a structural one, calling for a more coordinated approach, including at national level to meet the challenge[14]. The division of labour in the development of eResearch approaches between research workforce (academics, scientists), and a professional IT workforce is not only a structural challenge in terms of the "filling of the gap" in the expertise between them, but has also had an effect on communication channels. Thus AIMS makes the point that the most substantial communication about the Capability (in their case), came through professional IT circles rather than research communities and individuals.

ARCCSS tackles the question of the tension about where investment should lie as between a data-centric view versus an HPC-centric view. It must be said that this is evident in some of the commentary, but the point is made by this Centre that this is a pointless difference as all components are needed. The same response goes on to express concern about the way eResearch might be given leadership "at all levels" and this should not be seen as a cost –

> Wings on an aeroplane are not an expense; they are fundamental to the operation of the plane. E-research is fundamental to the capacity of Australian researchers to maintain international competitiveness and continue to drive discovery.

[14] See the discussion at Section 4.7 in the full Status Report, or under Section 5 in the Summary Report.

Finally, CSIRO takes the opportunity to offer a more sweeping view about a possible rethinking of the structural approach for future consideration. It is quoted here –

> During the past 8 years technologies and cultures have matured such that many institutions are now hosting their data and/or business services through offsite providers, and have become increasingly comfortable supporting technical interactions via remote collaboration environments rather than local staff presence. As a result it may be worth reviewing the model whereby independent and transient regional service providers in each state each deliver a comprehensive set of eResearch services to mostly localised clients. There is scope to reduce duplication by concentrating large data holdings and compute infrastructure to a smaller number of physical locations that operate on a truly national scale over the long term. This could then be complemented through a single distributed "virtual centre" of expertise for eResearch services, which provide skilled specialist support staff with a long-term home, a sustainable career path and an optimally connected peer community. It would be appropriate in this context to ensure that the capabilities delivered through NCRIS funded programs are now focus funded in the near term to be optimised through sustainable human capital investment over continued distributed technology propagation, as this is likely to be a defining difference between moderate and significant transformative *value from investments to date.*

The service providers had a number of distinct comments on structural issues as follows –

- ANDS reports that, if the national approach had not been taken, the structure of research in institutions would lead to suboptimal outcomes in the (necessarily) collaborative task of improving data management and sharing.
- AAF is concerned that although categorised as part of the NCRIS eResearch Capability, it has not actually been funded in that way.
- RDSI responded that an alternative would have been a non-federated approach through a single operator, but it is not clear whether that would have had the requisite buy-in and co-contribution.
- NCI responded that there has been a failure in general of universities to invest in HPC, relative for example to the UK.
- NeCTAR commented that the approach fostered high degree of collaboration within domain focussed proposals. However sufficient high quality skills and expertise was not available to the level needed for optimum establishment and deployment of the infrastructure. Furthermore the skills deficit will worsen as the funding cycle (and the projects it has enabled) completes.

Technology Developments or Deficits

Most feedback here is about development and its speed, rather than deficits. In some cases there is a view that the technology itself has been driven by some of the Australian investments, but in turn agencies such as Geoscience Australia consider that technology developments have transformed that body's research activities. The issue of balance of investment is also raised, including a view about a greater impact if it had been more oriented to peak facility programs (Go8).

The speed of change would indicate that future projects need to plan with greater agility, particularly in thinking about the mix between outsourced and in-house developed technology.

There is also some comment on how technology procurement looks from the point of view of industry. While from a procurer's point of view, it might well be that there is some danger of partial obsolescence if there is a long lead time, a vendor's alternative view is that the speed of technology cycles make tendering more ambiguous and uncertain, and so suboptimal.

It is worth making a point here about the extent of collaboration between industry, in this case vendors of high performance capability, and the organisations and bodies that are established to engage in the purchasing process. There is some evidence of variability in this from the point of view of vendors. In some cases, adversarial approaches to IT procurement, while they may satisfy a sense of professional competence and pride on the part of the procurer, are not necessarily helpful, at least in the context of a transaction involving high degrees of complexity and leading edge technology. The professional expertise

and knowledge developed at NCI has been informally commended from both procurer and supplier perspectives in this regard.

As commented elsewhere, both the AAF and RDSI are components of the Capability where the relevant technologies were, at initiation, and are still, maturing. In the case of the AAF access requirements continue to be complex. In the case of RDSI the comment is made that-

> At the start of the project, the technology to achieve the current RDSI outcome was neither available nor known. The RDSI project has needed (in partnership with others) to develop these capabilities in order for the project goals to be achieved.

NeCTAR makes a significant and lengthy commentary on technology. It is quoted here in full.

> The emergence of the open-source OpenStack cloud computing fabric arrived "just-in-time" for NeCTAR with the establishment of the OpenStack foundation shortly before the first deployment of the pilot node of the NeCTAR Research Cloud. In the early days of the Research Cloud deployment, deploying, managing and operating OpenStack entailed heroic efforts by a small, talented and committed technical team. While NeCTAR's selection of OpenStack as the research cloud platform in April 2011 may have been considered a "brave choice", the subsequent development of OpenStack and the high levels of commitment to the product from major commercial vendors and developers has vindicated that choice. Today, OpenStack is considerably more mature technology which has been supporting a stable, reliable and scalable cloud computing platform fit-for-purpose for the Australian research community.
>
> The emergence of cloud computing has had a profound impact on the commercial software industry, impacting on software architecture and design methodologies. Cloud computing has also fostered increased ability for innovation in software NeCTAR has been engaging with the (e)research software development community to support improved understanding of the benefits of cloud-enabled software design and architecture. This has included funding for Research Application Migration activities and support for training roadshows and the development of researcher-focussed documentation.
>
> Key research domains internationally are embracing cloud platforms for deployment of the next generation of the research software infrastructure. In many respects, NeCTAR has established a leadership position for cloud-based research e-infrastructure.

APPENDIX FOUR

References

Works Cited in the Text

2010 NCRIS Evaluation Report

Australian Government. Department of Innovation Industry, Science and Research, National Collaborative research Infrastructure Strategy, Evaluation Report, June 2010.

2011 Roadmap

Australian Government. Department of Innovation Industry, Science and Research, 2011 Strategic Roadmap for Australian Research Infrastructure, September 2011.

Advancing Canada's digital infrastructure

Canada. Advancing Canada's digital infrastructure, The International DI Context: Digital changes everything…, [prepared for DI Summit 2014], http://digitalleadership.ca/wp-content/uploads/2014/01/The-International-DI-Context.pdf

AeRIC Annual report

Australian eResearch Infrastructure Council, Annual Report 2012

The Australian Research Data Infrastructure Strategy

Australian Government, NCRIS, The Australian Research Data Infrastructure Strategy, 2014

Cyberinfrastructure Framework for 21st Century Science and Engineering

[United States] National Science Foundation, Cyberinfrastructure Framework for 21st Century Science and Engineering, 2012

[http://www.nsf.gov/cise/aci/cif21/CIF21Vision2012current.pdf]

The Emerging Governance of E-Infrastructure

Barjak, Franz, et al, The Emerging Governance of E-Infrastructure, in, *Journal of Computer Mediated Communication*, 18, (2013), pp 113-136.

ESFRI Roadmap

European Commission, Directorate-General for Research and Innovation, Assessing the Projects on the ESFRI roadmap, (A high level expert group report), 2013

The eResearch Coordinating Committee Final Report-2006

Australian Government, Department of Education, Science and Training, & Department of Communications, Information Technology and the Arts, e-research Coordinating Committee, An Australian e-Research Strategy

Tildesley Report

United Kingdom. Department for Business, Innovation and Skills, A Strategic Vision for UK e-Infrastructure, 2011

[https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/32499/12-517-strategic-vision-for-uk-e-infrastructure.pdf]

Other Works Consulted

Australian Government, Department of Industry Innovation, Science, Research and Tertiary Education, 2012 National Research Investment Plan

Australian Government, Department of Innovation, Industry, Science and Research, Strategic Roadmap for Australian Research Infrastructure, August 2008

Australian Government, NCRIS Committee, Strategic Roadmap, 2005

NCI, (unpublished discussion paper*),* Advancing Australian Research and Innovation Capabilities: The Case for a National e-Infrastructure and High-End Computing Strategy, August 2014