

Submission

2016 National Research Infrastructure Roadmap Capability Issues Paper

Name	Professor Hugh Durrant-Whyte ¹ Dr Bradley Evans ²
Title/role	¹ Director of Sydney Centre for Translational Data Science ¹ Director of Sydney Informatics ² Associate Director of Sydney Informatics ² Director of TERN Ecosystem Modelling and Scaling Infrastructure
Organisation	The University of Sydney

Questions

Question 1: Are there other capability areas that should be considered?

There is a gap in the provision for Translational Data Science, which we discuss in the context of Environmental and Earth Observation in questions 18,19 and 20.

In summary, we propose an (Australian) Translational Data Science Facility ((A)TDSF) as a newly funded NCRIS capability, its key role will be the;

- Transformation of discovery in NCRIS science driven by modern machine learning and data science tools. Specifically across physical, life and social sciences.
- Enabling scientists to connect with all Australian and international data and develop exploration and discovery applications.
- What we are not going to do is build or support hardware.

We view data and discovery in three layers:

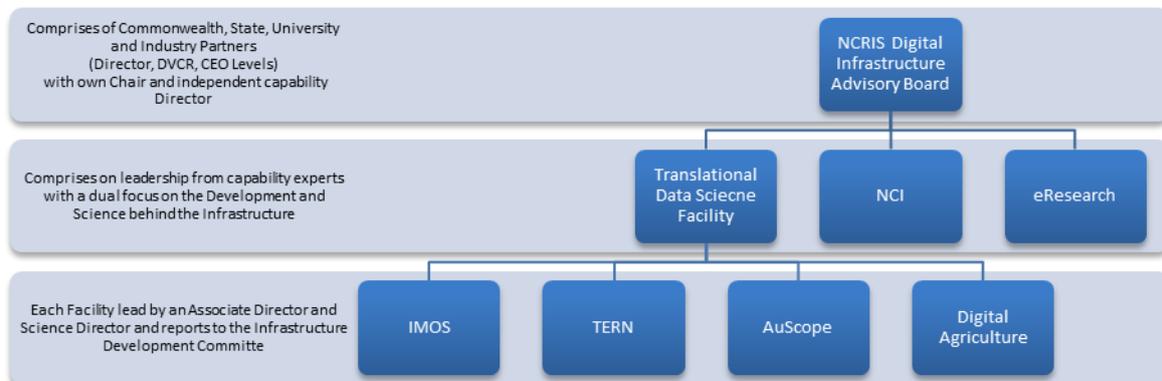
- (IaaS) Data and computing support (AWS, NCI, RDS, ANDS, etc)
- (PaaS) Cloud and support for networked data/computation (NECTA, NCI, etc)
- (KaaS) Delivery of data science tools for scientific discovery – in eg TERN, IMOS, AuScope, etc.

TDSF focus is on delivering the last of these three – the main route to generating outcomes for science and society.

Question 2: Are these governance characteristics appropriate and are there other factors that should be considered for optimal governance for national research infrastructure.

Governance needs to span across the Commonwealth, State, University and Industry. The wide spread investment in peak computing (NCI) and eResearch requires this for ensuring the infrastructure is developed in response to community/ user need and then attracts the necessary

uptake and co-investment required to keep it operating sustainably. A proposed model for governance is as follows;



Question 3: Should national research infrastructure investment assist with access to international facilities?

Absolutely. Consolidation of national and international investment is key to reducing the cost of infrastructure delivery in Australia. Whilst sometimes, it is better to build our own capability, it is often more cost efficient to co-invest with international partners. Strategically, this can practically mean reaching agreement amongst national networks to pool resources and support international activities. For example, the development of Cube satellite technologies requires a range of infrastructure (launch, ground communications) which extend beyond the capacity of Australia’s research infrastructure. Working with partners such as the European Space Agency, NASA, JAXA and others we can deliver better outcomes to Australian researchers through investing in gaps in the international investments- leveraging our role within the larger project.

Question 4: What are the conditions or scenarios where access to international facilities should be prioritised over developing national facilities?

The choice for when to invest in national infrastructure over international collaboration should follow the following ranked criteria;

- (1) **National security and data security:** Some information is of key national significance, for example, the linking of yield information to crop area before trading is of significant importance to not only market speculators, but also food security. Protecting this information until it can be made completely publically available is paramount.
- (2) **Return on investment:** Key criteria for investment are (a) life cycle cost of the infrastructure (b) industry alternatives to building bespoke solutions (c) barriers of entry for users (4) outcomes of stakeholders (beyond cost savings). For example, whether to build a national cloud as a service infrastructure over using Amazon.

- (3) **Capacity/ innovation building:** The choice to develop infrastructure with a known life-span may include the utility of the potential capacity development, or innovation potential as a criteria where (2) is not significant.

Question 5: Should research workforce skills be considered a research infrastructure issue?

Absolutely, one of the key lessons of NCRIS so far is that uptake is always limited by user capacity. Education and training, which yields change, can take several years to achieve a paradigm shift. Whilst infrastructure must stay ahead of the “workforce skills” curve, it must also integrate more completely with education and training.

Question 6: How can national research infrastructure assist in training and skills development?

Through the provision of mandate and resources to engage with STEM skill education across all levels of the education sector. Conversely, through incentivising the creation of curriculum which works interactively with NCRIS to provide services useful for education. Whilst these should be based on real data and services, some provision should be imagined for Education. One barrier to this has been the lack of long-term commitment to services. Educators, like infrastructure providers do not have the budget to re-invent education material every year as technology changes or services are made unavailable. Some provision for longevity of service should be built into the system. We should envisage a future where educators look to NCRIS for access to tools, information and services for providing “real life” tools for educating tomorrow’s researchers.

Question 7: What responsibility should research institutions have in supporting the development of infrastructure ready researchers and technical specialists?

Similar to the mandates and resources mentioned in Question 6, research institutions should co-invest in the research infrastructure. For example, the approach of the University of Sydney is where possible to develop internal capabilities, which either merge or parallel external capabilities. These provide the necessary capacity building to deliver on NCRIS infrastructure and similarly the skills to help make the best use of it.

Question 8: What principles should be applied for access to national research infrastructure, and are there situations when these should not apply?

Open access, with attribution, should be the default method of access. However, given the targeted nature of many NCRIS investments, it may be appropriate to solicit basic information - without coupling any restriction to access – to assist in delivering a more user focussed experience. For example, requiring login via a national service or a validated email address with an associated account linking to basic information such as organisation, location and voluntary information such as research interests and project details.

This approach should of course not apply to access which is of concern to national security.

Question 9: What should the criteria and funding arrangements for defunding or decommissioning look like?

The following criteria should be considered;

1. The term of the total investment, as human resources are institutionally bound to enterprise agreements covering their rights- of particular concern for majority NCRIS funded capabilities. For example, severance payments.
2. Certain capabilities require alternative arrangements before decommissioning is announced. For example, long term data storage or compute allocations underpinning NCRIS research.
3. Likely community response to decommissioning.

Question 10: What financing models should the Government consider to support investment in national research infrastructure?

Funding should be relevant to the size of the industry and its importance to national GDP, the financial and outcome related benefits of the infrastructure (relevant to it not existing) and underpinned by our national scientific strategy. Commonwealth agency priorities for investment should not be wholly allowed to drive NCRIS investment. NCRIS should consult widely with all layers of Government, Industry and focus on University co-investment in the infrastructure because they provide the critical link to (a) Postgraduate research resource – necessary for insuring the best possible research infrastructure and (b) have to consider a wide range of researcher requirements and not just those which return highest profits or operational objectives.

Question 11: When should capabilities be expected to address standard and accreditation requirements?

Resource permitting, at all times.

Question 12: Are there international or global models that represent best practice for national research infrastructure that could be considered?

None identified at the time of submission, echoing how this is first for Australia and the world.

Question 13: In considering whole of life investment including decommissioning or defunding for national research infrastructure are there examples domestic or international that should be examined?

In the USA, their National Ecological Observatory Network (NEON) have written and will soon publish a book¹ on the lessons learned of the consequences of decommissioning infrastructure and provides details which include international examples.

¹ Chabbi, A and Loescher H (eds) (in press) Terrestrial Ecosystem Research Infrastructures: Challenges, New Developments and Perspectives. CRC Press/ Taylor and Francis Group

Question 14: Are there alternative financing options, including international models that the Government could consider to support investment in national research infrastructure?

NCRIS could consider, under a capability expert, leveraging international investment in this space for delivering a Digital Agriculture infrastructure funding model that supports the basic connectivity to ecosystem data (climate, ecology, weather and soil) to underpin further research and product development and promote competition, which will deliver higher productivity and a more sustainable Agricultural industry.

Health and Medical Sciences

Question 15: Are the identified emerging directions and research infrastructure capabilities for Health and Medical Sciences right? Are there any missing or additional needed?

NA

Question 16: Are there any international research infrastructure collaborations or emerging projects that Australia should engage in over the next ten years and beyond?

NA

Question 17: Is there anything else that needs to be included or considered in the 2016 Roadmap for the Health and Medical Sciences capability area?

NA

Environment and Natural Resource Management

Question 18: Are the identified emerging directions and research infrastructure capabilities for Environment and Natural Resource Management right? Are there any missing or additional needed?

The gap in Australia's digital research portfolio is an NCRIS capability focused on the challenges of transforming data into knowledge. The current investment into Australia's data and compute infrastructure has seen the proliferation of two very strong hardware and platform layers. The first, lead by peak data and compute capabilities the NCI and Pawsey, the Infrastructure as a Service layer (IaaS). The second the Platform as a Service Layer (PaaS) lead by ANDS (metdata), NeCTAR (cloud compute) and RDS (Data) has created a distributed network of infrastructure for localised or regional projects. These projects have highlighted the strength in integrating hardware and software. However, we believe more can be done in this space.

Building on IaaS and PaaS (in other words – together with) we propose a Knowledge as a Service (KaaS) we call (Australian) Translational Data Science Facility ((A)TDSF) . Make no mistake, this is not another hardware layer, this is a software knowledge layer designed to integrate with existing NCRIS infrastructure and work across data collecting capabilities TERN, IMOS, AusCope, BioPlatforms,

Groundwater, ... and the newly proposed Digital Agriculture and Space capabilities. Critically, the new facility will integrate with the peak and eResearch capabilities.

The (Australian) Translational Data Science Facility ((A)TDSF) will;

- **Work together with NCRIS capabilities** to develop the software necessary to not only use data more easily, but transfer this into tangible knowledge. This is not research, it's a research infrastructure to deliver the most efficient way of turning information into knowledge and giving the users the choice: raw data or knowledge. Liaising with Australia's world leaders in all scientific disciplines will remain a key part of all translational activities.
- **Breaking down barriers between researchers and the knowledge they seek.** One of the key criticisms of NCRIS capabilities is the barriers of entry to accessing data. We will address that by building new software infrastructure that considers the full range of use cases and is not biased by scientific interest as previous attempts have been.
- **What we are not going to do is build or support hardware.** This is not another hardware layer. We have sat down with the NCI and eResearch community and said, this is our expertise, we want to work together with you. This is what we can do. We do of course understand and use the technology everyday, however, we are software engineers and scientists. Not hardware and networking specialists.
- **National in scope, a world first.** Whilst we are the first University based investment of its kind in Australia, we recognise that other Universities and Scientific Agencies will gladly join our national facility. We welcome collaboration, after all, we are proposing services for all to consume. If they don't want to collaborate, they will still be considered welcomed and valued users of our infrastructure.

Question 19: Are there any international research infrastructure collaborations or emerging projects that Australia should engage in over the next ten years and beyond?

No, as far as we can tell, we are world leaders in this space. It is important to note that we expect this capability will attract significant international attention over coming years and again, as with the NCI, TERN and IMOS, we expect to be world leading.

Question 20: Is there anything else that needs to be included or considered in the 2016 Roadmap for the Environment and Natural Resource Management capability area?

We recommend a restructuring of the peak data and compute capabilities and the establishment of an overarching body to manage funding into initiatives such as our own. We would also expect our approach to be central to any over-arching body seeking to focus investment across IaaS, PaaS and KaaS. We believe we are capable of leading such a coordination activity given the nature of our expertise is to work across disciplines and layers of technology.

Advanced Physics, Chemistry, Mathematics and Materials

Question 21: Are the identified emerging directions and research infrastructure capabilities for Advanced Physics, Chemistry, Mathematics and Materials right? Are there any missing or additional needed?

NA

Question 22: Are there any international research infrastructure collaborations or emerging projects that Australia should engage in over the next ten years and beyond?

NA

Question 23: Is there anything else that needs to be included or considered in the 2016 Roadmap for the Advanced Physics, Chemistry, Mathematics and Materials capability area?

NA

Understanding Cultures and Communities

Question 24: Are the identified emerging directions and research infrastructure capabilities for Understanding Cultures and Communities right? Are there any missing or additional needed?

NA

Question 25: Are there any international research infrastructure collaborations or emerging projects that Australia should engage in over the next ten years and beyond?

NA

Question 26: Is there anything else that needs to be included or considered in the 2016 Roadmap for the Understanding Cultures and Communities capability area?

NA

National Security

Question 27: Are the identified emerging directions and research infrastructure capabilities for National Security right? Are there any missing or additional needed?

NA

Question 28: Are there any international research infrastructure collaborations or emerging projects that Australia should engage in over the next ten years and beyond?

NA

Question 29: Is there anything else that needs to be included or considered in the 2016 Roadmap for the National Security capability area?

NA

Underpinning Research Infrastructure

Question 30: Are the identified emerging directions and research infrastructure capabilities for Underpinning Research Infrastructure right? Are there any missing or additional needed?

NA

Question 31: Are there any international research infrastructure collaborations or emerging projects that Australia should engage in over the next ten years and beyond?

NA

Question 32: Is there anything else that needs to be included or considered in the 2016 Roadmap for the Underpinning Research Infrastructure capability area?

NA

Data for Research and Discoverability

Question 33 Are the identified emerging directions and research infrastructure capabilities for Data for Research and Discoverability right? Are there any missing or additional needed?

The gaps identified in this proposal represent a significant interest in the development of data for research into Agriculture, TERN, IMOS, AuScope and food production.

Question 34: Are there any international research infrastructure collaborations or emerging projects that Australia should engage in over the next ten years and beyond?

1. The Global Soil Map project (<http://www.globalsoilmap.net/>) is working on projects to extend the current mapping strategy down to higher resolutions; the current international target is 30m. Coupling the soil algorithms with remote sensing such as Landsat and Sentinel.
2. ARC ITTC Food Safety in Fresh Produce has a number of projects looking at the relationship between challenges faced by our food industry and climate, transport and safety concerns.
3. The European Unions Horizon 2020 programme will spawn a number for projects such as "Food watch from space", "Monitoring Soil Moisture for Optimised Irrigation", "Sustainable Forestry Management" and more. <http://www.copernicus.eu/main/agriculture-forestry-and-fisheries>
4. POLARIS: A 30-meter probabilistic soil series. (University of Nebraska, Princeton, The University of Sydney, USA National Soil Survey Center).

Question 35: Is there anything else that needs to be included or considered in the 2016 Roadmap for the Data for Research and Discoverability capability area?

We recommend a Digital Agriculture capability be established with a specific mandate to integrate with existing peak data and computational infrastructure such as the NCI and Pawsey and the wider eResearch community (ANDS, Nectar, RDS) and Environmental Capabilities (TERN, AusCope and IMOS).