Submission
2016 National Research Infrastructure Roadmap Capability Issues Paper

Name	**Virtual Laboratories Advisory Group**
Organisation	**NeCTAR**

General Comments:

The Virtual Laboratories Advisory Group provides high level guidance to the NeCTAR Virtual Laboratory Program. Comprised primarily of senior researchers from the Virtual Laboratory community, the Virtual Laboratory Advisory Group is responsible for providing recommendations and guidance on areas including:

- Strategic planning for the Virtual Laboratory program
- Responding to and influencing of broader agendas including NCRIS
- Promotion of the Virtual Laboratory program
- Engagement with the research community

A list of members is attached at the end of this submission.

Nectar provides an online infrastructure that supports researchers to *connect* with colleagues in Australia and around the world, allowing them to *collaborate* and share ideas and research outcomes, which will ultimately *contribute* to our collective knowledge and make a significant impact on our society. Nectar Virtual Laboratories are rich domain-oriented online environments that draw together research data, models, analysis tools and workflows to support collaborative research across institutional and discipline boundaries. They are built and led by the Australian research sector and are used nationally and internationally by the research community. These programs have been very successful in engaging the research community, industry and international partners. The Virtual Laboratories are utilised by over 10,000 researchers from a wide variety of disciplinary backgrounds.

Virtual Laboratories are exemplars of the ideal research data system proposed in the National Research Infrastructure Roadmap Capability, combining compute, data resources, reference data, authentication and access control, provenance, publishing and of course analysis into a set of community-based collaborative environments. They provide benefits for researchers in high tool availability and accessibility, data and model sharing, convergence of training and practice environments, economies of scale with regard to compute and data resources, and leveraging of benefits of other capabilities (eInfrastructure particularly). They also facilitate international participation and leadership in best practice initiatives such as the Human Brain Project, ELIXIR https://www.elixir-europe.org/, NIH Data Commons https://datascience.nih.gov/commons and the Cancer Genome Atlas

http://www.cancergenomicscloud.org/. Virtual Laboratories significantly enhance opportunities for disciplines and are central components in the Australian Research Data System.

It is critical that infrastructure for research data continues to support sophisticated digital research methods as well as the innovative application and development of tools. With increased longevity of availability of data, improved data curation, data and analysis tools are now closely aligned, with increasing emphasis now needed on digital research methods and the software infrastructure that supports this.

Questions

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

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

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

It is important that national research infrastructure investments in data assist with access to international data facilities, to facilitate increased international linkage of data. Linked data is critical in many research communities, including the humanities.

At another level, it is valuable to ensure that access to international facilities does not prevent Australian infrastructure investments from continuing to advance. Major European and US investment towards specialised online digital infrastructure, such as Virtual Laboratories (or "Digital Gateways") will naturally outweigh Australian investment. For example, a sizable proportion of the €1B Human Brain Project (HBP) is devoted to developing sophisticated online platforms for neuroscientists to perform simulation experiments, and to access important data collections. It is natural that Australian researchers will increasingly want to use these internationally developed platforms through collaboration or other means. Using overseas digital environments will undoubtedly provide Australian researchers with value. However, it is essential that Australian researchers become the "Virtual Laboratory builders" as much as the "Virtual Laboratory users". There is significant benefit to being involved in the design, development and innovation that occurs when developing digital research platforms. The knowledge generated is immediately transferable to the knowledge and innovation economy.

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

Usually scale and sensitivity are the defining factors. For example, in astronomy a consortium of EU countries can build a telescope that can see significantly more than anything a single country alone (e.g. Australia) could build. If you want to be competitive you need to join such consortiums (or invest 10x more).

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

Research workforce skills have been repeatedly highlighted in sector review, and should be considered a research infrastructure issue. To enable Australia to continue to lead internationally in research data and collaboration infrastructure, this is a key area that needs development. Essentially there needs to be recognition and career pathways for this workforce. The model of short term "interim" funding means that many highly technical staff at the leading edge of virtual laboratory infrastructures are forced away from academia into other areas. Ideally there should be 5-7 year funding available to ensure that the best staff can have guarantees of continuing employment.

The humanities approach to lifting research workforce skills through "user centred" information and data design has had significant effect on addressing lower levels of data skills in that sector. For example, humanities data has traditionally been developed by individual researchers or within highly specific disciplines, and skills in data interoperability were not always a priority despite the evident gains. Research infrastructure is by definition connective, and scaled developments such as HuNI directly address these traditional humanities practices as a skills deficit.

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

There are a myriad of ways in which training and skills development can be enhanced, including systematic case studies showing that this is happening but could be done much more. For example, national facilities could be encouraged to offer intern programs, or partner with universities to allow PhDs to perform some of their research away from campus and under their guidance.

It is also worth noting that the national infrastructure is already heavily used for training and skills development. Over 500 Master's level students have been trained at the University of Melbourne on Cloud Computing (using NeCTAR) and using it for a multitude of big data analysis problems. Many of these students have since moved into significant positions in industry (Google, etc) or been employed in eResearch groups directly to continue their work. It would be extremely difficult to imagine how this would be achieved without facilities such as NeCTAR.

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

Research institutions should continue to facilitate development of key research skills, including relevant technical expertise as well as actively participating in discoverability infrastructure. See above comment also - hands-on training on the Cloud and big data technologies (CouchDB, Hadoop, Spark etc) is essential for many disciplines. Many universities have recognised this and put their own programs into place, e.g. degrees in Data Science, which cover several of these issues. There is a clear need to teach the next generation of infrastructure ready researchers and technical specialists about the major technologies influencing current research efforts.

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

Access should be as open as is sensibly possible (modulo privacy issues, IP, etc). The value derived from the data, and hence return on investment, is maximised when the data is exposed to the maximum number of researchers.

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

All projects should ideally have measurable KPIs and be assessed against them. These could include metrics such as number of users; number of publications; number of spin off grants/companies; measurable improvements identified by community (couldn't live without it etc). Where possible, defunding/decommissioning could include consideration of the transferring of the cost of operations to industry/private partners, including a supported pathway to achieve this.

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

A key problem currently is that funding allocations are typically small and timelines short. This makes long term infrastructure planning (and in particular, retaining good staff) difficult. The long-term stability of national research infrastructure should be an important consideration in the funding model.

It is also the case that ways to combine State-based investments and federal investments could/should be explored.

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

The problem is that many research areas don't have agreed standards, or that the standards are only marginally useful for the goals the researchers are trying to achieve. And often no one can agree on what the standard should be. Only when the standards in a research discipline are mature enough should they be enforced.

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

In the humanities Australia is well ahead of the curve at the national level in terms of digital infrastructures, with Virtual Laboratories like HuNI and Alveo providing best practice examples. However, more can be achieved in terms terms of international interoperability.

The model taken up by the astronomy community in the early/mid-2000s, as recommended by that decade's Decadal Plan, has worked especially well. Here, the major universities and observatories set up a not-for-profit called Astronomy Australia Limited (AAL). AAL identifies and represents our national interests and works with the Government to realise them. They manage a number of our large

infrastructure projects and collaborations, as well as seeding and helping develop smaller projects for the community.

NeCTAR has globally leading areas that have galvanised the international community. This is also quantifiable, e.g. the endoVL project has over 100 hospitals globally using the platform and systematic analysis of global publications related to adrenal tumours has clearly demonstrated that the project has shaped global research, i.e. major global publications increasingly from users of the endoVL platform.

It is noted that there are many cases where the model of infrastructure provision has been less successful and infrastructure efforts subsequently decommissioned. The UK e-Science efforts and European Grid efforts are prime examples. There are few identifiable capabilities that persist from these initiatives.

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?***

International examples include the National e-Infrastructure Service, which aims to facilitate UK research by providing access to a broad range of computational and data based resources (http://www.ngs.ac.uk/nes). The European Grids for e-Science (EGEE http://cern.ch/egee) is another example. In the former case it could be argued that this was too generic (primarily HPC focused and not targeted to specific user communities), whilst in the latter it was too focused on the high energy physics community and their very unique challenges/demands and complexities of the underlying software systems.

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

The National Health and Medical Research Council (NHMRC)-European Union (EU) Collaborative Research Grants scheme is one model, which aims to provide assistance to Australian researchers to participate in multinational research collaborative projects with international researchers that have been selected for funding under the European Commission's Horizon 2020.

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?***

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

There are numerous international research infrastructure collaborations that Australia should engage with, including projects on cancer genome, and multiple personalised medicine activities (H2020 grants etc). Importantly there has to be a mechanism to engage (as an equal partner). At present only the NHMRC provides funding (typically up to $100k per annum) to be involved in successful EU projects. There should be a review of this across other non-medical research areas, e.g. ARC.

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?

Ensuring that the work is aligned with international efforts should also be considered. Co-funding for involvement in consortia should be better aligned (including H2020, NIH, others).

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?

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

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?

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?

The text should refer to “Astronomy and astrophysics”, not “Astronomy and cosmology”. Cosmology is just a sub-field of astronomy.

In terms of emerging directions, gamma ray bursts (GRBs), the epoch of reionization, and gravitational wave astronomy should be mentioned.

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

It’s already mentioned briefly, but joining ESO should have more emphasis. The stability and long term value of a multi-government supported multi-infrastructure model cannot be understated.

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?

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?

The issues paper asserts that collecting institutions are “vital research infrastructure”, but this is not necessarily how they function for researchers who are only one of their user groups. The national institutions are information providers with users from across the entire community, and their funding

and services are intended toward much broader goals. The specific ways in which these collections can be defined as a "national digital humanities capability" for research requires very careful thinking and extended consultation with researchers themselves. Consequently, the relationship of the national collecting institutions to the national research infrastructure that has been developed and delivered by researchers themselves is an important one which deserves fuller and more careful consideration than the current document proposes.

There is also an increasing focus on cities with a multitude of efforts exploring opening up data sets related to urban/situational contexts by local, State and Federal government departments. There is a need to consider cities as research entities in their own right since they provide a living platform of research and data: health, transport, housing etc. There are many groups and efforts globally exploring smart cities for example and optimised ways to reuse data for more intelligent decisions and city planning. It is noted that such activities are directly relevant to the public - an aspect that has been increasingly emphasised in what the NCRIS capabilities need to demonstrate as value for money for infrastructure funding.

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

The emergence of networked data structures in the humanities (of which HuNI has been a leading proponent) is suggestive of many future collaborations. HuNI has been contacted by many European potential partners (we are only hampered by lack of access to resources to ensure these relationships are developed). Current open offers for collaboration exist with:

- http://histograph.eu/
- http://www.europeana.eu/portal/en
- http://wiki.dbpedia.org/

HuNI's approach to vernacular (unstructured) data ecologies is also being used as a "model of interest" by the commercial publisher Elsevier.

The LAPPS Grid (USA) and CLARIN (EU) are two important consortiums and initiatives for language and speech data. The Alveo Virtual Laboratory is collaborating with both and ensuring interoperability wherever possible.

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?***

The question of how best to reboot Trove as a research infrastructure (rather than as an information provider) is not really addressed by the existing document. The National Library of Australia should also be included for consideration.

National Security

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

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

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

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?

The work of the Australian Access Federation has been key in developing an effective national Access and Authentication system that simplifies access to infrastructure. It would be useful to support international single-sign on via the AAF and international federations.

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

Considerable savings could be made in astronomy by joining ESO and gaining access to a suite of next-generation telescopes, underpinned by the long-term stability and funding model of the EU (notwithstanding the Brexit!).

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

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?

To participate in large-scale international endeavours, It is important that Australia provides the collaborative funding mechanisms required to participate in international programs, such as Horizon 2020 and its projects such as HBP. This informatics work doesn't have a natural home in one of the existing Underpinning Research Infrastructure projects, and therefore the opportunity to leverage funding to promote closer international collaboration is lost.

A single entity for Data for Research and Discoverability would create a better opportunity to provide access to resources across multiple NCRIS capabilities (NeCTAR, RDS and ANDS). It would also assist in enabling capture of all of the informatics work being done for a given community (i.e. Characterisation) to establish international connection opportunities. The timing is important: Australia is currently well placed to develop a strong national and international user base of online digital research platforms, based on national experience and expertise built through NeCTAR Virtual Laboratory program, and the Research Data Services A1.x program.

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

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?

There needs to be strong emphasis on Virtual Laboratories, cloud-based storage and computation. Australian researchers utilise a network of interacting digital platforms for their research. To achieve the vision of creating future research applications that exploit innovative research methods, infrastructure must also emphasise the provision of improved access to methods and data, cloud-based storage and computation. The Nectar Virtual Laboratories and Research Cloud programs have provided nationally significant infrastructure impact to a number of research communities, with demand clearly remaining for more opportunities. Both Virtual Laboratories and Research Clouds including commercial cloud, international science clouds, Australian research cloud, storage and computation, and cloud enabled research software and data, enable rapid access and interoperation at national scale.

Other comments

List of Virtual Laboratory Advisory Group members:

Virtual Laboratory	Advisory Group Representative
Genomics Virtual Laboratory	A/Professor Andrew Lonie Director, VLSCI
MARVL Marine Virtual Laboratory	Dr Roger Proctor eMII Director, Integrated Marine Observing System, IMOS
Virtual Geophysics Laboratory	Dr Carina Kemp Senior Geophysicist Geoscience Australia
Climate and Weather Science Laboratory	Dr Aurel Moise Senior Research Scientist Bureau of Meteorology
Characterisation Virtual Laboratory	Professor James Whisstock Director of the Australian Research Council Centre of Excellence in Advanced Molecular Imaging, Scientific Director of EMBL-Australia, Monash University
All Sky Virtual Observatory	Professor Darren Croton Professor of Astrophysics Swinburne University of Technology
Humanities Networked Infrastructure	Professor Deb Verhoeven Chair in Media and Communication Deakin University
Industrial Ecology Lab	Dr Tommy Wiedmann Associate Professor of Sustainability Research UNSW
Biodiversity and Climate Change Virtual Laboratory	Professor Brendan Mackey Director, Griffith Climate Change Response Program, Griffith University
Endocrine Genomics	Professor Richard Sinnott Director of eResearch University of Melbourne
Alveo	Associate Professor Steve Cassidy Department of Computing Macquarie University
Virtual Hazard Impact & Risk Laboratory	Dr Jane Sexton Senior Scientist Geoscience Australia