

Submission

2016 National Research Infrastructure Roadmap Capability Issues Paper

Name	Dr Peter May
Title/role	Head of Research
Organisation	Bureau of Meteorology

Health and Medical Sciences

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

The Bureau considers that most broad areas of capability of concern are covered in the paper, although suggest that more consideration needs to be given to computational science and high performance computing in particular. Computational science underpins advances in each of the focus areas and is an area where skills and workforce need to be developed including algorithms, software development and machine learning for advanced architectures and exascale futures. Future high performance computing platforms capable of exascale-level performance are required for advances in environmental science and other areas in order to undertake numerical experiments at increasingly high resolution to resolve key processes.

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

The Bureau considers that the governance characteristics cited are appropriate to helping prioritise research infrastructure. Frameworks for accountability are important especially to ensure clarity of coordination of investment and the use of shared resources. For example, the Bureau considers that strengthened governance around the use of high performance computing for research purposes may help optimise availability of resources for research infrastructure at the National Computing Infrastructure (NCI) and the Pawsey Centre. In both cases a strategic approach will assist the centres and stakeholders plan for increasing demands, the challenges of increased data volumes and capabilities needed to make best use of the available resources.

Governance models need to capture the connections between management of research and operational infrastructure as operational infrastructure often provides data for research purposes as well. This will deliver efficiencies in operations (economic value) and streamline new operational system innovation.

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

Yes: the Bureau concurs with the dimensions identified in the Roadmap. The Bureau's own experience is that support from a national research body and/or research capability carries weight when negotiating enabling partnerships and collaboration with international agencies, such as National Weather Services. Other examples of direct interest to the Bureau where national investment in research infrastructure capability would enable and yield considerable benefits include high performance computing, the management and analysis of large data sets and partnerships such as IMOS that have strong international links. Further, bi-directional access to

research high-performance computing and observational facilities enables closer collaboration on the science that in turn directly translates to operational outcomes.

The Bureau considers that this work on national research infrastructure has considerable potential to build confidence in Australia as a valued contributor to international collaboration, including through the IPCC and the Earth System Grid Federation at the National Centre for Atmospheric Research. That in turn will generate considerable returns to Australia's own research capability. While it might be considered that there may be a lag in delivery of the benefit through engagement with international facilities, the benefit gained is likely to be more significant and more enduring.

Last—and this is also relevant to question 4 below—access to international facilities could be parleyed through quid pro quo arrangements, effectively building complementary arrangements and capabilities. For example, Australia may not be able to even afford to contribute directly to large scale space-based research infrastructure (for example, for Earth observations from space and other space-based applications), but can contribute through other means of direct import to the research and project delivery (for example, through innovative instrument design, calibration and validation cases and systems, data processing and evaluation, and the early uptake of research outcomes). Some of these quid pro quo measures may require leverage from other research infrastructure capabilities.

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

International facilities are particularly important where there are opportunities to partner in facilities that are beyond our current funding or technical capability, but where there is a clear benefit in access to the facilities. Access to remotely sensed data from space is one area where both operational and research satellites provide immense value. Data from these systems are a key part of some existing facilities (e.g. IMOS).

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

The Bureau considers that research infrastructure cannot be separated from the skills, capabilities and resource bases needed to; operate, support and derive value from it. If Australia is to extract the highest value from our investment in research infrastructure, both the design of the infrastructure to address relevant national research objectives, the design of experiments and conduct of research using the infrastructure should be considered in parallel as well as the skills base to efficiently operate and make full use of the infrastructure.

In this context, the Review should also consider how funding for research infrastructure capability is structured. Rather than focussing primarily on capital, funding should reflect a more holistic approach to the capability that the Australian Government is investing in. We suggest that proposals should including sufficient operational funding to ensure the workforce skills are present to support the operations, and upgrades, to the capability during its lifespan. We also suggest that sound investment practices such as estimating total cost of operations, and a full life-cycle plan, which should include the workforce needed to support the infrastructure be applied. Such costing and appraisals do not exclude provision of support and skills sets from the private sector.

However, as noted above, to gain the best value from the investment, consideration of the readiness and availability of research skills able to make best use of the capability needs to be considered. That will be in a number of forms, from researchers ready to use the infrastructure immediately, to establishment of research programmes and research training needed to exploit the capability. Gaps may present opportunities to deepen international collaboration through invitations to international research groups and consortia, and to build private sector capability—including as with the Hartree Centre in the United Kingdom, small and medium enterprises. In the latter circumstances, research support—operators, engineers and data and computational scientists—may be needed to assist such opportunities.

In terms of developing that research, and research support, skills base, giving national researchers the opportunity to work with international infrastructure will expose them to new ideas, and increase their ability to contribute to national research goals.

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

Research infrastructure is necessary for training and skills development by providing the environments for both training of graduates and re-training of staff in new and emerging technologies. Training and skills development will not occur unless the infrastructure is in place and this requires a conscious investment choice. Opportunities could be offered for formal training such as traineeships, PhD scholarships and for taking advantage of the Australian Governments Innovation agenda. The Review may also wish to consider the value of ‘e-research’ capabilities—networked access to data and high-level compute, fast application development, and ready collaboration and training (including with industry)—as a specific national research capability in its own right.

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

The Bureau considers that institutions should provide input to national plans, such as this one, and that those plans should include contributions to and support of the human capability needed to support the infrastructure as well as derive value from the research outcomes. That said, it is a community effort. Research institutions should partner with academic institutions (where the two are not the same), operational agencies, other government entities and industry to support the development of skills required to meet changing user demands. While formal teaching is the province of universities, it remains the case that it is underpinned by research capabilities. As such there is a deep connection between research, research infrastructure, the skills and research training generated by universities, and value derived by industry and the community in terms of skills and capabilities of graduates.

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

The Bureau agrees with the accessibility considerations identified in the paper. It considers that overarching research merit (perhaps as per National Science and Research Priorities) is a priority driver, but that it needs to be balanced with the need for diversity of research areas and to increase research maturity in emerging areas. The latter would include areas where private

investment is likely to be low or negligible or that provides a basis for basic research that can be used and leveraged by government and business to create value.

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

The Bureau considers that a whole-of-life approach should be taken to any infrastructure and research infrastructure is no exception. Funding for end-of-life decommissioning and any upgrades should be factored into the total asset cost at the proposal stage, as well as all costs associated with the operation of the infrastructure. Both the total cost of operations and return on investment should be assessed as part of decision-making.

It is not unreasonable that a decision might be taken to defund/decommission existing infrastructure, to allow new priorities to be addressed, but adequate time and funds must be given to allow those dependent on that facility to source new funding or to plan their exit strategy. The timing required will depend on the nature of the infrastructure. For example, if funds are to be withdrawn from infrastructure that involves field-deployed measurement technology the researchers can weigh up the value of seeking out additional time in the field versus setting aside funds for recovery of equipment, make-good of site and shutdown of support systems, some of which may have been contracted well in advance. Unless an arrangement was agreed in the beginning, the national infrastructure funding should cover the costs of recovery, make-good and shutdown. Similarly, development or procurement of bespoke technologies may have a long lead time, and timing and funds will need to address possible exit clauses from procurement contracts.

In the event that the infrastructure has potential for transitioning from research to operational status (for example, a trial observing network that aligns well with ongoing operational requirements), early discussion and funding of the transition to operations process may facilitate continued value being delivered. However, continued operations would depend on availability of operational funding. Consideration could be given to continuation of full or partial operating costs, depending on the extent to which the infrastructure will continue to provide a foundation for new research infrastructure to be developed and/or to operate with synergy.

Another important consideration for effective transition of relevant research infrastructure to operations (versus decommissioning) is the adoption of data collection/reporting/coding etc standards from the beginning. Clearly some research infrastructure serves no ongoing priority need once the research objective is met and/or becomes lower priority. However, by starting with the possible end in mind (be it decommissioning or transition to ongoing operations), adopting a whole-of-life funding model, and making informed choices about data models, unexpected costs will be minimised.

Last, a transition to operations involves quite a different mindset, governance and funding arrangements than may have been appropriate for a purely or even partially research-oriented capability. Any value assessment would also need to consider stage or life-cycle and potential replacement for a service that may have many dependencies (including network and communications). Costs are likely to be higher than for a research capability (for example, increased robustness, energy use, wear and tear), though it is possible that these may also realise a

lower overall costs through scaling to many users. It is also worth considering that technological change may see first:

- 1) research capabilities transition to operations more quickly than in the past;
- 2) there is more likely to be private sector alternatives more quickly on the market, once a research capability has made that transition; and
- 3) there is a commensurate increasing risk of fast obsolescence of capabilities that were or remain in a research capacity, without increasing updates.

There are examples where facilities decide to keep infrastructure beyond end of life, but as a partner there is a dis-benefit that needs to be recognised in terms of increased costs of maintenance. Running infrastructure beyond its end of life should be addressed as a separate business case for power, facilities; people etc. The Governing boards should make specific determinations of the cost benefit of keeping something going beyond its planned end-of-life.

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

Research infrastructure will not be delivered without capital funding as well as operating funds—that is intrinsic to the assessment of the total cost of operations. Operational funding alone might work for existing infrastructure (but would not meet all decommissioning costs) but new infrastructure that cannot be leased, other than perhaps those things that can exist 'in the cloud', will require capital. A range of current approaches to this are undertaken at the various NCRIS facilities and we think this is appropriate.

Financing models should be based on areas of national interest that build resilience and lift capability. While a grants system allows investment in new research areas, there should also be allowance for building a foundation of research infrastructure that can be leveraged for multiple uses. High-performance computation is one such enabling capability. For example, the use of the NCI as a research platform, used by multiple entities and uplifting this to include a network of high performance computing both nationally and internationally.

Funding for the NCI has been asset based with much of the operating costs funded by the partners. Therefore, in this model the partners need to be convinced that there is value in continuing their contributions. NCI and Pawsey, in particular, need to be recognised as central assets in a national strategy for high performance computing – recognising the relatively fast life-cycle of computers compared to some other national facilities, which then informs the funding cycle.

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

As mentioned above, if there is any possible goal relating to transition of research infrastructure into operations, then early consideration of relevant standards, for example, measurement and data structures, will greatly facilitate and reduce the costs of the transition process. The production of research data from the national facilities has to meet format and discoverability standards as well as quality metrics.

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

The Australian Community Climate and Earth System Simulator (ACCESS) model development is managed through a partnership between the Bureau, CSIRO and Universities, and includes international engagement and governance through the Unified Model partnership agreement with the UK Met Office and national meteorological agencies. This recognises the magnitude of the asset that is ACCESS and the need for national and international capability to maintain the system as a world class modelling system. Such software should be treated as an asset just as hardware is.

The Unified Model development is managed via a partner contractual agreement between international partners, including joint governance and funding commitments, which is an effective operating model.

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?

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

For some applications, an alternative to funding traditional research infrastructure is the provision of funding of IT-as-a-Service (ITaaS) to allow for a scalable use of research environments and virtual laboratories in the cloud such as NeCTAR Cloud used for the Climate and Weather Science Laboratory. This may mean a move away from traditional asset based funding towards funding for operational expenses. However, high performance (super) computing needs require significant capital investment, operational support and ongoing, frequent upgrades and are required for a wide range of assets and programs.

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?

Nil response.

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

Nil response.

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?

Nil response.

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 Bureau supports the main conclusions of this chapter and strongly supports the need for sustained research programs for ocean, atmospheric and terrestrial monitoring. In addition the section should discuss critical needs for high performance computing, storage and corresponding eResearch. This includes an "active data" approach to facilitate the interrogation and use of large data assets. The Bureau welcomes the explicit inclusion of the Australian Community Climate and Earth System Simulator (ACCESS) model in the discussion paper, including the need for infrastructure to facilitate the broader use of the ACCESS model in the research community. The ACCESS model suite represents a critical asset for meteorological and climate research and requires significant resources from the operational and research sector and leverages off significant international collaboration. Leading weather and climate research requires such model infrastructure and this represents a need and an asset just as much as physical infrastructure. Without this as a tool for weather and climate modelling, there are substantial risks for planning and decision making in a wide variety of sectors. This is an asset that has an immediate path to significant societal and financial impact.

World competitive Earth system models are extremely complex with more than two million lines of code and are beyond the reach of individual institutions to build. The Bureau together with CSIRO and in partnership with the Universities have collaborated with the UK Meteorological Office, the US National Oceanic and Atmospheric Administration and other international partners including NIWA to build the ACCESS model suite over the past decade. This has resulted in significant outcomes with the Bureau's operational forecast modelling suite amongst the best in the world and ACCESS being among the best performing models in the Coupled Model Inter-comparison Project models for IPCC. However, the model infrastructure, including data input and managing, curating and analysing the vast amount of data output, as well as the ability to set up varied experiments etc. is limited because of the operational focus of the Bureau and CSIRO. In addition, it is useful to note that the speed of operation of models such as ACCESS can be optimised for particular platforms resulting in improved efficiency by a factor of two to three. This has been demonstrated at the NCI facility as part of a collaboration between NCI and Fujitsu. Modest investment in IT infrastructure located at a suitable research high performance computing and storage facility for the ACCESS model with a focus on research needs will greatly increase the utility of ACCESS for fundamental research and enable significant increases in our national weather and climate research capability. This infrastructure is quite separate to the operational needs of the Bureau with its focus on robustness and security prohibiting this flexibility in its operational environment.

In summary, the development of infrastructure in support of ACCESS requires an integrated platform of services, tools and data for researchers and government agencies to simulate, analyse and predict climate and weather phenomena at a national high performance computing facility such as NCI. The NeCTAR CWSLab (Climate and Weather Systems Laboratory) can be viewed as a proof of concept for this, but does not deliver the required flexibility, traceability and support for the community.

Models such as ACCESS are increasing in resolution and complexity as well as there being a growing need for ensembles of model output to explore variability and predictability. These research model suites also include high resolution capability for the ocean and atmosphere (in the hundreds of metres). Note that a doubling of resolution, while allowing increased fidelity, comes at the cost of about ten times the required computing resource. This combination is driving both the compute capability as well as storage and analysis technology. However the benefits are large with the promise of advanced warning of El Niño events, substantial improvement in tropical cyclone intensity understanding and forecasting, damaging flood forecasting, and bush fire risk all key research targets. The benefits of the research will flow through into operational outcomes through the Bureau representing a clear path to impact for the research enabled by investment in the ACCESS infrastructure.

This section should include a discussion on high performance computing and storage and capabilities for the curation and analysis of large data sets including through eResearch programs. The current NCI facility needs substantial upgrades to meet global standards for weather and climate model resolution and is behind the capability of the Bureau operational HPC. While noting the recent agility funding has been granted \$7 million and there is co-investment of \$7 million from the NCI collaboration, this is only an incremental increase in capability and will still leave the NCI machine lagging expected requirements – and critically the supercomputer will be at absolute end of life at 2018. To be world standard for HPC and to meet the community needs we require substantial increases in capacity, regular upgrade cycles and a plan for exascale computing within a decade. Exascale will be required by the environmental modelling community and there is clear demand for this in other fields as well.

This section should also touch on the benefits of HPC and high performance data, and complex models: enabling smart decisions for weather, climate, engineering design, insurance, safety, and the opportunity for economic benefit. In addition, the ability to manage, organise and analyse large data sets is a key capability: one that provides advice, analysis, and opportunity for economic benefit. The national high performance computing infrastructure needs to enable the production and analysis of immense data sets.

An emerging direction is environmental prediction includes transport of chemical species and air quality. This will require access to a much wider variety of data and the ability to use (assimilate) such data.

Another highly relevant emerging direction not mentioned is cities and built infrastructure. Integration of data (environmental, demographic, transport, industry, health etc.) from multiple sources, use of highly adaptive models to analyse and make short-term predictions, have the potential to increase the efficiency, productivity, environmental performance and safety of people, businesses and infrastructure in the urban environment. ACCESS development and appropriate HPC is required to meet the challenges implied by this need.

The paper supports the observations programs that are crucial and notes the importance of corresponding operational data such as the Bureau of Meteorology and Geoscience Australia collects that are also critical and often used resources for research. The research programs (IMOS and TERN) are crucial for ocean and terrestrial observations for the nation's research, as is the national Cape Grim Science Program contributing to national and global research and monitoring

outcomes in relation to atmospheric composition, but further investment is required to gain further research, policy and economic benefits

Under Section 6.3 in the Issues Paper, we agree with the need to build up national capability in sensors and sensor networks, with an important caveat on the importance of an integrated design approach based on the outcomes sought, to ensure that needs (research, national, economic etc), models (analysis, predictive, conceptual), other data sources (including existing/new fixed and remote-sensing networks) are considered in delivering value-for-money research outcomes. Standard formats, data discoverability and flexible access are key requirements. Note that data assimilation such as in the ACCESS model is the key to integrating a wide variety of data sources into a single optimised data set that is often far more useful than the individual data sources in isolation.

The paper states that 'access to international satellite-based remote sensing data' is a priority as yet to be addressed – the Bureau of Meteorology, CSIRO, GA and others coordinate their activities and have all individually established connections with international space agencies to access and use such data. The real priority is for a coordinated national approach to accessing, managing etc. these data, to build on the individual networks and connections, to avoid duplication, and to provide a nationally accessible repository for such data in support of national research priorities. Such an approach would also facilitate more effective integration with in situ observations, especially important in the boundary layer and terrestrial zone, where environmental monitoring is especially important.

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

Weather and climate research infrastructure are intrinsically international. The Australian community is closely engaged with the Unified Model partnership which is a cornerstone of the ACCESS model suite. The Australian contributions from the Bureau of Meteorology, CSIRO and the Universities is coordinated with efforts from the UK Meteorological Office, the Korean Meteorological Office, the Indian Centre for Medium range Weather Forecasting and New Zealand's National Institute for Water and Atmospheric Research.

Australian scientists, especially at the Bureau of Meteorology, are active contributors to the World Climate Research Programme and the World Weather Research Programme. These programmes bring together scientists from around the world and leverage research and operational infrastructure of national meteorological services and global research institutions. Australian scientists have leadership roles in the design and delivery of these international research programmes and continued engagement is essential.

Earth observations from space remains a key requirement for weather and climate research including the ocean and cryosphere, and Australia depends deeply on international collaborative projects and frameworks to enable access to much needed data, with quid pro quo via Australia's ground-based support (for example, the Bureau of Meteorology's satellite monitoring, data relay and turn-around-ranging facilities), calibration/validation sites and algorithm development.

As noted in the paper, that in addition to operational agencies, IMOS currently contributes to the Global Ocean Observing System and collaborates with similar institutes in the USA and other countries. Continued priority for IMOS as national research infrastructure is strongly supported.

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?

As discussed above, there is a critical need to include high performance computing and associated infrastructure and tools for curating, managing and analysing large data sets as well as the capability to run state of the art weather, ocean and climate models. The Bureau welcomes the explicit inclusion of the ACCESS model in the discussion paper.

The paper discusses the requirements for observations, but increasingly these data need to be integrated into a model analysis. This process, referred to as data assimilation, is a key capability for producing dynamically consistent continuous data sets from multiple data sources and data with significant gaps. Data assimilation is a key capability in weather, ocean, streamflow and seasonal forecast models, but it would be useful to highlight this as a national capability requirement.

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?

Nil response.

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

Nil response.

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?

Nil response.

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?

Note comment made under question 18 on urban environments and value of an integrated approach to environmental, culture and community data.

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

Nil response.

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?

Nil response.

National Security

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

A deep understanding of the natural environment is critical for national security, including emergency services.

Cybersecurity is a clear and emerging concern, especially in terms of code development and assurance, the management of big data, and in sensor rich networks, such as operated by the Bureau. Given the pace of the field, and the threat, we suggest that research be done collaboratively across government and industry.

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

Within the Cyber Security area, emerging international opportunities are difficult to acknowledge and track as research and innovation is tightly held by foreign governments and collaboration is not generally acknowledged outside secure networks.

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

Besides Cyber Security, perhaps consider peer-to-peer encryption technology.

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?

Computational sciences, skill development and workforce development were identified in the United States¹, United Kingdom², and European Union³ as emerging directions of improvement to

¹ Future Directions for NFS Advanced Computing Infrastructure to Support U.S. Science and Engineering in 2017-2020, <http://www.NAP.edu/21886>

² A Strategic Vision for UK e-Infrastructure, https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/32499/12-517-strategic-vision-for-uk-e-infrastructure.pdf

support future scientific and engineering development, as well as being critical to national competitiveness. New algorithm and software engineering development with advanced architectures and processors is an emerging direction in exascale computing and application design. Emerging directions in data analyses, machine learning and data-intensive computing are utilising new architectures, advanced processors and algorithms to analyse data collections resulting in new types of applications and services in science and engineering may be of use for some facilities.

Capabilities for computational science development to support future science and engineering, is visible in the current roadmap but are more prominent in international initiatives and venture capital investments.

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

Australia should continue its engagement in the Earth System Grid Federation (ESGF) over this period. Currently NCI hosts a node of the ESGF which facilitates ready access to a wealth of climate model output data, produced both locally and internationally, and adhering to strict data standards, to support climate impacts and adaptation analysis, and climate science more broadly. NCI and Australian researchers participate in international committees responsible for planning and overseeing the development of the ESGF and the form and scope of the model results available. However, this links to a gap in HPC and corresponding storage needs. The ESGF activity is currently unfunded.

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

A national strategy and initiative for advanced computing infrastructure and computational science capability needs to be considered, to ensure Australia's ongoing competitiveness in research and innovation. A review of the U.S. and European initiatives could contribute to the strategy and initiative such as the U.S. National Strategic Computing Initiative, U.S. NFS Advance Computing Infrastructure initiatives, U.S. DoE Advanced Scientific Computing Research (ASCR) program, the EU Framework Programme for Research and Innovation: HORIZON 2020 High Performance Computing.

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 paper acknowledges the importance of a nationally coordinated approach to research data and appears to address the main elements needed to support such an approach. Perhaps an added dimension that would be of value is a 'so what' element – an ability to also retain a

³ EU Framework Programme for Research and Innovation: HORIZON 2020 High Performance Computing, <https://ec.europa.eu/programmes/horizon2020/en/h2020-section/high-performance-computing-hpc>

coordinated approach to how the infrastructure has been used, link to research outcomes (papers, case studies etc.), value delivered, value-added information (products, services, data sets etc.).

A notable omission is the need to define, promulgate and support a standards-based approach to data discovery, access and re-use across research domains. This is critical for "active data" allowing an interrogation approach to data use. Significant progress has been made around Discovery through the efforts the Australian National Data Services, but access and re-use remains a major challenge within and across data domains. This becomes increasingly important if researchers and research infrastructure need to collaborate globally.

Refer also to earlier comments about satellite (and other) remotely sensed data and the need for a national approach not just to holding (maybe not all), discovering, accessing the data, but also for a coordinated approach to international relationships to support a comprehensive and collaborative National approach.

Emerging directions in data repository services, universal identifiers and programmatic data services are enabling new agile web-based application environments with access to digitized data and geospatial data collections. The NeCTAR Virtual Laboratory provides data and tools to researchers. There is a global trend in building these science gateways next to very large data archives, but Australia's investment is currently relatively small.

Whether we are talking 'big data' or 'small data', the ability to access readily and integrate freely will be the key to extracting future value, both through research and through delivered operational systems. This is where economic value lies and national research infrastructure is key to unlocking.

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

There are many initiatives that support data availability and exchange to support research such as the World Meteorology Organization's Commission for Climatology and Commission for Basic Systems on Big Data and Data Management and Data Rescue. International Data Centres represent 'centres of excellence' for collection of data to support research such as the US National Centre for Environmental Intelligence in the USA for climate data and for Global Precipitation Climatology Centre in Germany.

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?

The ability to work with streaming, unstructured data that is ingested from third parties should be considered as mentioned in our response to Question 27. This increases the need to define, promulgate and support a standards-based approach to data discovery, access and re-use across research domains.

Other comments

If you believe that there are issues not addressed in this Issues Paper or the associated questions, please provide your comments under this heading noting the overall 20 page limit of submissions.

The issues paper places emphasis on international drivers and collaborations however omits non-research sector national drivers. A number of emerging national policy drivers require consideration when designing the next phase of research data infrastructure investment, particularly from a data and discovery perspective (Section 12). These include but are not limited to activities supported by: the Australian Government Public Data Policy Statement, National Principles for Environmental Information, data.gov.au, the National Environmental Information Infrastructure (NEII), and the Platforms for Open Data initiative led by the Department of Prime Minister and Cabinet in collaboration with Data61. The current Productivity Commission Review into Data Availability and Use may also provide importance guidance for the research sector that warrants improved alignment and specific treatment in the roadmapping process given the commonality of objectives.