Submission
2016 National Research Infrastructure Roadmap Capability Issues Paper

Name	Prof Margaret Harding
Title/role	Deputy Vice-Chancellor, Research
Organisation	The Australian National University

ANU Summary of Key Issues

Key issues for consideration in the development of the National Research Infrastructure Roadmap are highlighted below. Given that the roadmap is a critical document that will inform the policy framework and level of Government investment, specific feedback is provided on policy issues and the case for increased investment, and discussion points raised in the face-to-face meeting with the panel.

Governance

Research infrastructure priorities and capabilities, by their nature, have complex, cross-portfolio implications, and require a ***governance structure*** that will allow the detailed technical and financial evaluation from a ***whole-of-Government perspective.***

- Investments from all Government Departments need to be identified and coordinated to maximise and leverage NCRIS funding from the Department of Education and Training.
- A review of expenditure on infrastructure across all Departments was commissioned as part of the previous Clark review and this data, if available, is clearly an important input to proposing new funding models.

Governance and accountability models for all NCRIS facilities is essential and should address issues identified in the ***reviews of governance in all NCRIS facilities*** conducted in 2014/15 with a focus on partnership models and transparency.

Demonstrating the Return on Investment from NCRIS

Correlation between our ***export revenue from international students – research rankings – research infrastructure, as a basis to justify increased NCRIS funding,*** is difficult to quantify, however, our international research visibility and scale is recognised as of major importance in attracting students to our country. So whilst a direct relationship is likely to be impossible the significance of the international student contribution to the Australian economy needs to be recognised.

Consideration of ***inputs to international rankings*** may assist. Specifically, AWRU inputs the following measures into rankings. This is not a huge data-set; a critical review of the use of NCRIS facilities in supporting the research published in Nature and Science, or facilitating the work of highly cited researchers, may be informative.

> ***Nature and Science papers (20%)*** *The number of papers published in Nature and Science between 2011 and 2015. To distinguish the order of author affiliation, a weight of 100% is assigned for corresponding author affiliation, 50% for first author affiliation (second author affiliation if the first author affiliation is the same as corresponding author affiliation), 25% for the next author affiliation, and 10% for other author affiliations. Only publications of 'Article' type is considered.*
>
> ***Highly Cited researchers (20%) -*** *The number of Highly Cited Researchers selected by Thomson Reuters. The Highly Cited Researchers list issued in December 2015 (2015 HCR List as of December 1 2015) was used for the calculation of HiCi indicator in ARWU 2016. Only the primary affiliations of Highly Cited Researchers are considered.*

Engagement with industry and external users (non-University) users was raised in all reviews of facilities in 2014/15. Involvement of the Industry Growth Centres to identify impediments or ignorance to the capabilities and the best ways to connect industry with facilities would be welcome by the higher education sector. Consideration should be given to providing funding to assist relevant NCRIS facilities to pump prime

and develop effective user arrangements with business; they could be supported, either direct to the NCRIS facilities themselves or through the Growth Centres. Such resources should be part of a national network for engagement with business.

It is essential to recognise that NCRIS facilities contribute to both fundamental and strategic research and different timelines and industry engagement strategies will need to be developed with different sectors.

International Access and Funding Levels

- Should be restricted to a limited number of large scale facilities that Australian researchers should access via an ***international membership or subscription model*** (telescopes, synchrotrons, Hadron collider); these facilities are of a scale or size that Australia would not be able to host or fund and should be supported but the underlying rational for NCRIS support must be the health and international competitiveness of the Australian research ecosystem.
- The ability to connect to, and access many ***key international data sets***, is a key theme in many of the specific capability submissions; directing NCRIS funding to access these data-sets, as opposed to Australian consortia being established to partner, would need to address specific criteria.
- Effective mechanisms are required to consider joining and proposing new major international co-funding initiatives on a rolling three-year time frame.

Funding Models

A tiered set of criteria for facilities funded by NCRIS to support the various capabilities is recommended that recognises ***different scales of investment, as well as different timeframes for investment*** e.g.

- *Landmark and rolling programme (>100M over 5 years)*
  Nationally and/or Internationally significant at a scale of investment that no institution can operate itself, and **will always** require a significant level of Federal Government funding; e.g. Telescopes; Australian Synchrotron; ANSTO Opal Reactor; CSIRO & Antarctic Division Vessels; National supercomputer capability (NCI and Pawsey).

- *Nationally Important and Time Limited investment (>5-100M over 5 years)*

- **Finite investment** required: time limited either 5, 7 or 10 year review (all with mid-term reviews).
- **Flexible funding models** that accommodate the different type of facilities and needs; e.g., initial up-front funding may be required, with tapered funding after 5 years and for example zero funding after 10 years requiring a business model to be built with partners to achieve financial sustainability.

- Initial capital investment over given timeframe (e.g. 3-5 years), aligned with long-term recurrent investment for ongoing operational and maintenance costs (there is no point in funding capital development without recurrent support)
- Review criteria and timeframes should be specified when funding is awarded that will allow hard decisions to be made regarding further investment or decommissioning, balanced against the need to ensure strong return on investment.
- At a minimum a refresh of the roadmap every 5 years is needed to recommend what facilities should be defunded and over what time frame.

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?

Under capability 5.3.1, we endorse the approach but describe a need for a networked and distributed imaging research infrastructure, especially fMRI. These are used in both psychology research and population health research to study many areas of cognitive function and cognitive decline. While there are diagnostic machines available, these are fully committed to clinical service and are not generally available for research purposes. An increase is needed in the number of human diagnostic imaging capabilities nationally if Australia is to position itself internationally.

Under 5.2.3, though sequencing capacity is not really an issue at the present it is critically important that it is kept up to date and refreshed in a field which is advancing very rapidly.

Under 5.2.5, data analysis and storage is crucially important for "omics" projects. Genomics now is able to generate whole mammalian sequences at a great rate and data analysis is becoming the bottleneck. Access for researchers to analytic pipelines through the National Computational Infrastructure for all Australian biologists doing genomic or high throughput sequencing will be critical. The software-based pipelines are as important as the hardware. Data storage and curation is also needed as the amount of data is escalating. Each sequence is Gb in size. As sequences are often re-analysed, storage is also essential.

Under 2.2.4, Biobanks are extremely valuable infrastructure. They store samples that may be up to 100 years old and have been collected and collated for decades. These need effective preservation and to be curated effectively to ensure they continue to be useful to current-day researchers (for example through targeted genome sequencing). For example, a collection of Indigenous and Islander bloods, collected over 50 years ago at ANU, would benefit from such analysis and significantly impact on Indigenous health issues.

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?

New effective and efficient pathways to therapeutic compound discovery and translation are essential. This would require large scale, high quality animal breeding and genome engineering facilities coupled with high-throughput therapeutic target and molecule assay design and screening facilities, supported by medicinal chemistry capability. This infrastructure will allow an increase in the number of therapeutic lead compounds produced in Australia by increasing our ability to develop animal models of human disease, screen for compounds that ameliorate the disease phenotype in animal models, and, when promising compounds are found, to use medicinal chemistry to mould these into drug-like compounds.

There are three major infrastructure needs to be addressed if Australia is to take full advantage of its strength in biomedical research.

1. Large animal houses and a small number of highly efficient CRISPR/Cas9 facilities.
2. To generate more lead compounds we need to increase significantly high throughput assay design and high throughput screening facilities in Australia.
3. High capacity medicinal chemists with time and the ability to manipulate the structures of dozens of lead compounds annually. While this may look like research, it is an infrastructural need as this is part of the pipeline of targets through to drugs and the conversion of screen outputs into drug candidates.

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?***

We support the emerging directions identified with the following clarifications.

A key infrastructure requirement is to ensure the effective curation, development, and analysis of long term datasets (which once lost, cannot be regained). This therefore does not relate to equipment but to data curation infrastructure.

The Australian economy and society are highly dependent on natural and environmental resources, across a wide range of sectors including agriculture, water, recreation and tourism. These and other sectors are utterly reliant on strong national capacities in research, data conversion and application of knowledge. While Australian environmental sciences (eg. FoR codes 0401, 0406, 0501, 0502, 0602) are in total very strong Australia has the potential to be a world leader in the creation, integration and application of knowledge to natural resource and environmental policy and management. To achieve this, dedicated and reliable infrastructure investment in the form of people, site-based ecosystem data, and models is needed. This investment needs to build from current long-term ecological research networks and existing landscape, water and climate data assets. Investment and integration across people, networks and equipment is needed so that we can evolve current capabilities into long-term ecological-economic research and long-term ecosystem risk assessment networks, which can in turn, inform State of Environment reporting.

Whilst integration refers to the adding together or amalgamation of data, what is really needed is synthesis capability. Synthesis refers to the deriving of meaning from the data to produce an outcome. Application, in this context, refers to using the data to answer specific questions. These components of synthesis and application require understanding of ecosystem processes, and are not just a generic technological capability. Some pertinent examples of the broad conceptual frameworks needed for the application and testing of 'data integration and modelling capability' are the United Nations System of Environmental-Economic Accounting, the IUCN Red List of Ecosystems, and the Sustainable Development Goals. Research infrastructure to enable applications of these frameworks, at a scale appropriate to meet the national need, is required.

Climate and water resources should be expanded to encompass sustainable use of natural resources under conditions of climate change and greater resource demand. The issues of data requirements and scientific evidence to inform decisions about resource use is common to many resources: water, carbon, soil and biodiversity. Knowledge of these are equally important to meet the identified Science and Research priorities.

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

Investment is increasingly needed to enable Australian researchers to support both the increasing global research agendas driven by, and national and international policy decisions demanded by, international frameworks and their domestic application including:

- The UN System of Environmental-Economic Accounts
- Tracking and reporting on the UN Sustainable Development Goals
- International Panel on Biodiversity and Ecosystem Services
- Reporting against evolving post-Paris Agreement climate policy commitments
- IUCN Ecosystem Risk Assessment framework
- International collaboration under the coordination of Future Earth

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?

An additional capability need is a system to use existing and future environmental data *to measure and forecast* ecosystem change to support Australia's domestic needs under the EPBC Act and meet international obligations through agreements, such as those under the UN Framework Convention on Climate Change, the Sustainable Development Goals and the Convention on Biological Diversity (to name a few).

The 'emergent direction' that is needed is whole-of-system thinking and the need for coordinated and coherent experimentation, adaptation, and mitigation strategies. Large-scale manipulative experiments in particular are needed. This type of research infrastructure is needed to drive breakthroughs in techniques for ecosystem restoration. The ecosystem services needed to maintain a prosperous Australia will depend on restoration of ecological function. Embedded within this should be a focus on highly interactive species that help restore ecosystem function e.g. ecosystem engineers, and keystone species. These will be critical to soil health, carbon sequestration and pest management at a time when we anticipate ecosystem shifts (i.e. collapse of current functions), increased frequency of extreme events, island displacements, and sea incursions (among several other environmentally driven challenges) over the next 50 years. The 2016 Roadmap must clearly inform this pathway through dedicated and appropriate infrastructure investment.

Missing or additional emerging directions include:

1. Transdisciplinary and translational research (i.e. social, health and ecosystem scientists working together) and the infrastructure needed to enable this.
2. Research infrastructure for measuring and managing uses of, and users of, environmental resources and social-ecological systems.

Advanced Physics, Chemistry, Mathematics and Materials

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

Access to international facilities by funded-membership is an important funding model for Australian researchers. Examples of missing large-scale capabilities include:

- National Ultra-High-Field Nuclear Magnetic Resonance Capability
- Virtual National Photovoltaic Laboratory to support cross institutional access to existing infrastructure and ongoing photovoltaic technology development in Australia
- Enhanced cryo electron microscopy and crystallography capability
- Foundry for the custom design and fabrication of microelectronic devices and test structures
- Co-funding for Australian partnership in an 8-metre-class telescope
- A unified Australian astronomy infrastructure organisation for managing, developing and operating major national/international capabilities in place of the current distributed model
- Support for the Stawell Underground Physics Laboratory

Many Australian researchers rely heavily on access to unique and complementary equipment and infrastructure around the world and we therefore strongly support NCRIS-funded access programs. The nature of the access arrangements will clearly vary from facility to facility but it is important that such schemes are flexible and responsive.

Examples of international facilities and infrastructure include:

- International Thermonuclear Experimental Reactor
- Synchrotrons (e.g. ILL Grenoble, Argonne, Pohang, Skuba)
- Ion Accelerator/ Nuclear Physics Facilities
- EuroBioImaging/Synchrotron
- Telescopes (e.g. the Giant Magellan Telescope and the Square Kilometre Array)

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?

Alignment of the ARC's LIEF scheme with NCRIS. There are many examples of LIEF-funded equipment being under-utilised due to a lack of appropriate technical support and infrastructure. This should be addressed by encouraging much stronger alignment between LIEF and NCRIS. This could also promote the alignment of new capabilities with identified new and emerging research directions.

There is also currently a gap in funding between LIEF (typically <$1M) and NCRIS (which is best used for major facilities costing >$10M). The $1M to $10M range includes both NCRIS-worthy national research infrastructure and equipment that is more appropriately dealt with by LIEF (items that are for limited purposes or have less than national reach, but are of high scientific merit). Increasing the effective upper range of funding for LIEF and having some overlap in scale between LIEF and NCRIS would address this issue.

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?

We support the capabilities identified in the Issues Paper, and particularly the Working Group's recognition that the Australian Data Archive and National Centre for Indigenous Genomics are national infrastructures that are currently unfunded under national arrangements. We also support the recognition of the need for "digitisation infrastructure", noting that there is a distinction to be made between the infrastructure required for digitisation and the funding for digitisation of physical objects, and suggest that clarification is required on which is the priority.

Further areas to be addressed:

1. **Born-digital content.** Given the increasing proliferation of born-digital data in both the humanities and social sciences (HASS) we are pleased that observational and social data is incorporated in the Issues Paper. However, it is limiting to consider such content under the topic of "urban settlements". There is a broader range of potential applications within this capability that should be considered, such as online and e-government, ageing and wellbeing and political interest groups.

2. **History and longevity**. The study of communities and cultures is often a story of incremental and long-term change. There is limited recognition within the Issues Paper of the need to support historical content in the long term. Researchers in HASS will often seek to analyse change in the medium to long-term and therefore there is a need to enable both the study of long-term change, and to enable existing content to be examined in new ways. This can include the need for preservation of cultural content in the long term, including those collections that are not within the national and state collecting institutions, but that form part of a broader distributed national collection. It also includes the need to train future researchers in the management of digital data, (e.g. formats, associated meta-data, bundling related items) and in the research uses of digital data (e.g. maximising usability, reliability and validity, visualising results).

3. **Public institutions.** The Issues Paper explicitly recognises the role of the national and cultural institutions as an important resource for HASS researchers. We support the need for investment in these institutions to enable access to historical and cultural resources for research purposes. Research data resources developed within the public administration systems are institutional resources of relevance to this capability. For many social science disciplines (eg economics, public policy and education), these resources are arguably even more important than the institutional resources identified above. An exemplar here is the set of longitudinal survey datasets established by the Department of Social Services and supported through the National Centre for Longitudinal Data. One of these, the Household Income and Labour Dynamics in Australia (HILDA) study, has now been used in over 750 journal articles over its 15 year history. Similarly administrative data from government and other sources is of increasing importance for the study of long-term social and economic research and policy questions. This is a point reflected in the "Health and Medical Science Capability" (under "5.2.5 National health and medical data capability") but is not considered in this capability.

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

Australia is rapidly falling behind the rest of the world in research infrastructure for HASS. Other countries, particularly in the European Union, have established well-coordinated and resourced infrastructures across several HASS disciplines through the operation of coordinated national centres established with the support of government, and then coordinated through EU Framework funding for research infrastructure to establish

pan-European infrastructures. Effective international engagement in the next five years is essential to the health of HASS disciplines.

There are several key European infrastructures and national or multi-national programs that we would particularly recommend as relevant to this capability:

1. Social sciences
    a. Consortium of European Social Science Data Archives
    b. The European Social Survey
    c. Survey of Health, Ageing and Retirement in Europe
    d. Synergies for Europe's Research Infrastructures in the Social Sciences
    e. Cohorts and Longitudinal Studies Enhancement Resources
    f. International Public Use Microdata Samples

2. Humanities and Arts
    a. Digital Research Infrastructure for the Arts and Humanities
    b. Europeana Collections
    c. The Digital Panopticon

Australia has established similar infrastructures in many of the above areas, and there is often collaboration of Australian facilities with these European partners (such as the Australian Data Archive, PARADISEC, and the Digital Panopticon). The Australian infrastructures are funded through either institutional arrangements or one-off funding from ARC LIEF and other programs. This has resulted in limited potential for scalability without substantial additional resources, and risks to long term sustainability where institutional or soft-money funding is no longer available. Effective integration with LIEF in the NCRIS Roadmap is essential.

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?

There is a need for infrastructure which will promote and support cultural change and capacity building, to help researchers escape from the constraints of the desktop and commodity software, and bring these fields into the era of big data and data science. This will also need to include an emphasis on data provenance and integrity, and collaboration, data sharing and re-use, consistent with the FAIR principles (Findable, Accessible, Interoperable, Reusable) developed by the FORCE11 group. A first step would be a detailed roadmap and draft standards specifically for the HASS sector to support this capability.

We would also make additional points on the provision of infrastructure within this capability:

1. **Leveraging existing facilities.** The Issues Paper highlights the need for establishing a number of new capabilities, which we support. However, where possible, these capabilities should be enabled through the development and expansion of existing infrastructures within Australian institutions that already provide services and facilities to their disciplines. Some of these existing facilities are already recognised within the Issues Paper but there are others that could also fulfil capability needs with appropriate resourcing – e.g. the Australian Data Archive and PARADISEC – thus providing integration and scalability.

2. **Integration and standardisation.** There is a need for supporting the integration of facilities, through the use of interoperable and shared data and metadata standards. Several of the HASS disciplines that support this capability have well-developed models for data sharing and interoperability based on common metadata standards (such as the Data Documentation Initiative, in the social survey community, and the Text Encoding Initiative in the digital humanities and linguistics communities. These standards can and should form the basis for sharing resources between services providing this capability, as they provide a mechanism for both machine-to-machine and human-readable data sharing. These standards also serve to support the requirements of the digital repatriation capability

outlined in Section 8.1.2, by documenting the content of collections in consistent ways and enabling access and ownership rights to be asserted in relation to those collections.

Alongside the development of common standards is the need for integration of access to resources across facilities. The identification in the Issues Paper of the need for development of a national digitisation facility provides significant opportunities for both the digital humanities and social sciences (such as history and demography), but the digitisation infrastructure needs to be embedded within a framework for the sharing and accessibility of both the digitised resources and the infrastructure (hardware, software, networks and storage) that provide the digitisation processes. Better discovery systems for the infrastructure that exists, to enable researchers to find and use the digitisation facilities that are envisioned, will be critical to the success of such a capability.

The provision of such a discovery infrastructure also raises the potential for processing pipelines, with varying levels of automation and scaling of infrastructure, to be developed. Such projects would benefit from the development of digitisation facilities for the conversion of physical objects into digital formats. However it should be noted that there is additional benefit to be achieved where the digitisation is placed into a larger framework of related services, including data capture, data and object storage, and documentation of object relations and provenance of derived objects. Such related services also have benefit in disciplines outside this capability, such as the management of physical collections in earth sciences.

3. **Shared capability needs.** Our final additional recommendation is the need to consider the possible services available from other capabilities for this capability. Two examples include:

    - As noted above, public policy and social science researchers make significant use of administrative data and other government surveys in their academic and policy research. Access to data linkage and integration services such as those provides by the PHRN (mentioned in the Medical and Health capability) are also a need in the social sciences for the analysis of social policy outcomes
    - Access to secure facilities (such as the SURE system provided by PHRN) may also have uses in this capability. Access to restricted content is a corollary of the data linkage services described above, and might also be of benefit in the provision of access to confidential qualitative materials such as video and images.

National Security

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

We identified a number of gaps in the conceptualisation of national security in the Issues Paper, addressed below. Though detailed, the Issues Paper does not articulate the level of coordination required to address a critical national research endeavour. National security should be regarded as a 'whole of nation' endeavour, reflecting the complex relationships between security, national interest and societal wellbeing. It must reflect systemic risks and consider linkages across food, energy, water and environmental security as well as the traditional concerns of national security. These myriad interdependencies between national security challenges and opportunities warrant new research infrastructure investment to support Australia's security policy and operational development, but also to support Australia's key security role in the Indo-Pacific region. This is an area where a whole of government approach to infrastructure becomes critical.

Australia has significant national security research infrastructure capabilities for operational needs (particularly within the Defence agencies). However, no such infrastructure exists at a national level to situate operational capability development in the broader strategic context. For example, when mapping between NRIC areas and the National Science and Research Priorities on page 56, national security is missing in the matrix for food, soil and water, energy, environmental change and health. Cyber security is present, and also contributes to energy security in relation to, for example, electricity and gas infrastructure. In Section 9 on National Security, biosecurity, cyber security and water security are specifically addressed, but not energy security. We support the Issues Paper's assessment that cyber security is an emerging direction for broader, deeper and multidisciplinary research of national importance, and a critical area for research infrastructure development.

The Issues Paper notes existing infrastructure in the areas of Animal Health, Phenomics and life sciences genomics, but infectious disease genomics is not mentioned as a current resource or potential new infrastructure. The emerging directions listed in the Issues Paper do not take into account the global health security threats from infectious agents that affect human populations, many of which cross sectoral boundaries with animals and plants. We need to strengthen our preparedness for response and intelligence, both nationally and within the region.

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

There are a range of discussions occurring in international fora (such as the United Nations) and international organisations (such as NATO) that will likely see opportunities for future international research infrastructure collaborations on national security challenges. The roadmap should aim to be as agile as possible in the national security arena to enable Australia to contribute/participate in such opportunities, where it makes strategic and policy sense to do so.

One aspect of biosecurity that is not covered is the large-scale sequence data collections on pathogenic organisms that are currently being developed internationally that enable tracking of disease across international borders. Australia is not immune to incursions from pests and pathogens despite our high quality quarantine. An example of an international network that is beginning to track foodborne and zoonotic pathogens internationally is GenomeTrakr—an initiative to whole genome sequence pathogens from foods, animals and humans funded by the United States Food and Drugs Agency. European agencies are involved in multiple research initiatives into pathogen spread across countries, which rely on harmonised sampling and testing protocols, and a common European-wide database for sequences and meta-data. Given the global movement of foods, animals and humans, it is vital that Australian research facilities can contribute to these international initiatives.

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

We recommend the consideration of the following areas for infrastructure capability development for national security:

1. **An integrated national security assessment and modelling capability, focused on current and future threats and risks to Australia's national security**, would allow the fast-tracking strategic futures work including scenario development and the identification/discovery of positive and negative impacts across the spectrum of national security threats and risks, using a holistic and multidisciplinary approach. The capability would deliver an adaptive, evidence-based framework for national security policy and responses and would be the key research infrastructure to support the development of a national risk register able to consider systemic risks, cumulative risks and cascading failures.

2. **A federated cyber range capability**. Cyber ranges provide secure ICT environments to test (largely technological) cyber security responses, solutions and resilience. A handful of small-scale cyber ranges exist within Australian research institutions and the private sector. A federated capability would enable broad researcher access to such infrastructure at scale and the breadth of cyber security challenges and opportunities tested to be expanded. This would dramatically improve the quality and volume of cyber security research on technical and non-technical issues, as well as the intersection of issues.

3. **Network of animal and public health laboratories.** There is a compelling need to enhance Australia's preparedness for outbreaks of infectious diseases affecting humans, foods and animals, through enhanced capacity for whole genome sequencing. This is more about ensuring that there is a robust network. Such a network would facilitate public health response and associated research into incursion of new pathogens affecting the Australian environment and population.

4. **National hub for research and public health response**. Australia needs, as a priority, a human health laboratory designated as the national hub for research and public health response, equivalent to the Australian Animal Health Laboratory. Such a national facility would house a whole genome sequencing dataset that incorporates relevant meta-data, for capturing sequence data from pathogens of animal, food and human origin. This laboratory would provide access to public health services and researchers to national information from a variety of sectors to understand pathogen transmission and dynamics both within and external to Australia.

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 infrastructure that underpins research has overlapping missions with the infrastructure required to support data for research and discoverability (s. 11). Separating the two capabilities runs the serious risk of repeating the mistakes of the past. For example, there are lessons to be learnt through the administration of NeCTAR and RDS investments in software and data storage. These capabilities will need to operate in harmony, to have co-investment, and overlapping advisory structures to govern the communities using them, to ensure users are able to get information in and out of the platforms according to their needs.

The components of eResearch Infrastructure, High Performance Computing (HPC), High Capacity Networks (HCN), and Access and Authentication (A&A), within this capability should be treated together with the goal of providing one integrated national e-infrastructure environment (comprising a supercomputer, a high performance private cloud for data-intensive computation, fast large scale storage, technical and domain expertise etc). The fusion of big compute, big data and the expertise required to stitch it together is generally referred to as "Advanced Computing". We therefore group our comments relating to these items together under this heading.

1. **Advanced Computing Services**. A National Advanced Computing Service is needed to provide infrastructure of a scale and performance, and embody high levels of expertise, to:
   - Provide Australia with an internationally competitive, high-performance compute/data production service, the combined capability of which will rank in the world's top 20 facilities;
   - Ensure Australian researchers are connected to each other and to the outside world using the very latest networking technologies;
   - Offer best-practice, converged/integrated computational and data services (including curated national collections), shaped to meet the needs of Australia's research and innovation system, and made available through both merit and priority access mechanisms;
   - Sustain a critical mass of expertise that is respected nationally and internationally for the quality and the innovation of both the operational services and infrastructure integration; the research-facing service portfolio scoped to deliver transformative outcomes; and excellence of its contributions in computational and data science;
   - Serve as a national hub of expertise that can work with industry, government and universities to equip the existing and future workforce with the knowledge and skills they require to exploit advanced computing; and
   - Have the remit to represent Australia in international peak computing and data collaborations/forums, e.g., www.exascale.org.

   NCI and the Pawsey Centre provide the building blocks for achieving the above. There are, however, two major impediments:

   i. *Lack of long-term funding*. The lack of long term funding makes it hard to make long term commitments to hold important data sets or to engage in long term projects with other advanced computing facilities around the world.

   ii. *NCRIS funding model that requires all recurrent costs to be met through co-investment.* The lack of recurrent funding means that co-investment is largely focused on basic operations, with limited funding available to support the specialists needed to assist with the creation of next generation software. This is in stark contrast to overseas where significant expertise exists inside national centres and national laboratories. It is also becoming an increasingly critical issue, since performance gains are no longer evolving automatically from advances in the technology, but require the redevelopment and optimisation of software to exploit the

latent performance inherent in the technology. Moreover, there is also an increasing need for large scale applications to consider not just application performance but also application energy usage.

2. **Geospatial Systems**. Geospatial infrastructure is becoming increasingly important with the expected development of, for example, driverless cars, automated agricultural machinery etc. The infrastructure takes the form of ground-based receivers that track signals transmitted by satellites operated by international space agencies and/or governments. Position accuracy can be enhanced through the application of adjustments to the signals received by user equipment through communication of corrections over some broad-scale communications network. These corrections are derived from the ground-based geospatial infrastructure such as that installed by AuScope and the State Government survey departments.

   The same ground-based equipment can also be used for high accuracy scientific studies of crustal deformation, seismic hazard and even quantifying changes in water resources. In a direct sense, quantification of the strain accumulation across the Australian continent will be enhanced through ongoing measurement of positions. In addition, the definition of the reference frame (the basis of the global coordinate system used in scientific research) depends on geospatial infrastructure, which feeds into the accuracy with which satellite orbits can be estimated. This subsequently improves the accuracy of estimates of sea level rise and changes in water storage, linking directly to climate and water issues (s. 6.1.2) as well as water security issues (9.2.3).

   The identified need for "skilled personnel to develop algorithms and tools to fully exploit the data for the global satellite community" noted in section 6.3 also applies to the requirements for water security in Section 9.2.3 and could also contribute to underpinning research infrastructure.

3. **Neutron and X-Ray Scattering.** The Opal reactor at Lucas Heights and the Australian Synchrotron are of significant interest to researchers at ANU. New beamlines at both facilities, for example for circular dichroism spectroscopy should be considered and would support work on protein structures and novel materials.

4. **Digitisation.** This is an area of increasing importance and interest and the facility will often need to be linked to advanced computing services that provide data storage and analysis capabilities. Digitisation impacts on many disciplines from biology through physics to environmental studies and the arts and humanities. For example, the National Virtual Core Library was one of the funded components of AuScope, one of the NCRIS capability areas. It involved performing visible and infrared spectroscopy scans of a vast quantity of rock cores that were stored in an inaccessible way by State Government geological survey organisations. The intention was to unlock the information stored in the rock samples and make it available in digital form to researchers. The program was a great success, with over 2300 drill cores having been scanned (over 630,000 m of rock) since the commencement in the original NCRIS AuScope program. This demonstrates the immense value that investing in a national digitisation program can provide. More such programmes are required.

5. **Tier-2 Computing.** The Issues Paper refers to Tier 2 computing as having a focus on a particular research community or the needs of a particular institution. The usual definition of Tiers 1 and 2 relates to scale and performance, with Tier 1 typically being beyond the capacity of an institution to procure and sustain, while Tier 2 is within the remit of a single organisation or a community, e.g., the data handling back-end of a major instrument such as a synchrotron.

   We suggest the Expert Group needs to question whether there exists an ongoing market failure surrounding Tier 2 computing and storage that necessitates continued Commonwealth investment. Tier 2 computing and storage is largely a commodity these days, within the reach of any institution

which can acquire as little or as much infrastructure as it requires, dependent on its budget, or as a service through commercial cloud providers. In the absence of this being addressed, there is an ongoing danger of cost shifting of institutional responsibilities. If Commonwealth funds are to be invested, then efficiency through aggregation must be a hallmark, as must the clear intention of the funding to transition or accelerate a change in the data system. Co-investment will be a powerful instrument through which to perceive and demonstrate the value of any investment.

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

1. **International Exascale Project.** Major challenges exist in high end computing to reach the next level of performance (exascale) with systems that can process the volume of data delivered from the square kilometre array telescope, or that can run a complete simulation of the human brain. Challenges exist on multiple fronts. The US, Europe and Japan have developed plans to tackle these issues. Australia should look to develop a close partnership with at least one of these international initiatives.

2. **International Reactor and Synchrotron Facilities** for example, ILL in Grenoble.

3. **Space Missions**. On the time scales of 10 years and beyond, Australia's potential involvement in future space missions should be considered. While the Australian Space Utilisation Policy states that Australia will not launch our own satellites, it highlights the considerable involvement and contribution that Australia can make through the development of payloads for international missions and the algorithms required to analyse the data and calibrate/validate instrument performance. ANU has potential interest in this space through the Advanced Instrumentation Technology Centre but also through direct involvement in specific missions. For example, the ANU expertise and involvement in developing laser ranging instrumentation for the GRACE Follow-On mission is highly suited to the requirements for the GRACE II mission, currently a Tier 2/3 mission in NASA's Science Plan (2014).

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 future cannot be dictated by the past. As discussed above, the infrastructure required to support data for research and discoverability will need to connect advanced computing (i.e. the fusion of big data and expertise) with the intelligent management of that data (i.e. that which has been performed by ANDS and RDSI thus far). The relationship between advanced computing and data management will form a layer of closely integrated services and capabilities:

1. **Data Platform services**
   - Education: Research communities who wish to advance their research using advanced computing will need to better understand how to use these services.
   - Aggregation: the current spread across six platforms within Australia is an inefficient model. Reducing these to a few sites would produce economies of scale in their management and capabilities, and would release funding to be able to be used for Data Management services.

2. **Data Management services.** The key framework would ideally be guided by the FAIR Data Principles (i.e. Findable, Accessible, Interoperable and Reusable) developed by the FORCE11 initiative. To date, the ANDS project has been focusing on the Findable (i.e. Discoverable) aspect of this framework, although it would benefit from a common definition of what constitutes the term 'data' itself, given that some research communities do not currently define their research outputs in this way, e.g. photography.

   However, the other principles are yet to be approached from a national level:
   - Accessible: this rests on a fundamental assumption of standardised and free protocols to enable universal implementation and access.
   - Interoperable: this requires a linked, open data model, including common ontologies and data frameworks.
   - Reusable: this principle will ensure that the data will maintain its validity over time for research use by adhering to agreed standards and protocols for both data and metadata.

   There will also be shared features to these services:
   - Both sets of services need to be designed by the communities that will be using them, to ensure the access model is appropriate and clearly defines for users whom they need to engage with when wanting to consume services.
   - Defining these communities will require a degree of sophistication and understanding of the appropriate levels of aggregation. For example, HASS have overlapping but not entirely shared needs of these services, but are commonly grouped together for other purposes.
   - There will be a maturity variance between research communities in terms of their familiarity with available or supportive services, and how they might use them within their research.
   - A catalogue of services is essential, showing end-to-end processes. This will require a discovery phase to establish the pipeline of activity to service. This will also serve to raise visibility of available services.
   - A provenance model for linking and tracking movement of data through these services will be of increasing importance as well (e.g. the collaborative work between NCI and CSIRO).
   - Funding could spread along the NCRIS (i.e. community/nation/university) spectrum, dictated by proximity to service.
   - In addition, data capabilities need to continue to be developed in researchers, in particular postgraduates and early career academics. These capabilities need to include data curation and preservation and well as metadata to ensure that data is managed well and discoverable.

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

- Alignment with the work of Force11, in particular international guidelines on data citation and the FAIR principles would optimise the benefit from research data and engagement on an international scale.
- Within particular disciplines, there may be significant benefit in engaging with the various EU-funded "European Research Infrastructure Consortiums" funded under the Horizon2020 program. Examples of these for the HASS community are covered in the Understanding Communities and Cultures response.

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?

- Data which is not born digital needs to be considered. This is highly important when considering research in the HASS disciplines. Although the report talks about digitisation, it does not articulate a way to capture non-word-based artefacts, such as artworks, photographs, field notebooks, and increasingly video.
- There is concern that we don't currently have the skilled workforce in Australia to support some of these services, and that investment needs to occur to lift the local maturity in this area.
- Human resource capabilities need to be included into the user model. For example, educating researchers to incorporate appropriate metadata handling when initiating a project. There are new entrants to the industry every year, so this could include an induction program for PhD candidates; the use of existing staff within a University to reinforce this message; and the inclusion of a relevant clause in research contracts (currently being piloted within Horizon 2020).