**1. What is your occupation?**

- Researcher
- University Staff
- Business / Industry
- Government
- Higher Education Peak Body
- Industry Peak Body
- None of the Above

**2. Are you responding as an individual or an organisation?**

- Individual
- Organisation
- None of the above

**3. What organisation do you work for?**

National Committee for Data in Science (Australian Academy of Science)

**4. Are the recommendations appropriate to the current NRI environment?**

The Committee recommends that the challenge-based framework be refined to identify societal challenges.

Recommendation 3 is " Adopt a challenge framework to support NRI planning and investment". The Committee endorses this approach which has been successfully used for example by:

- the United Nations <https://sdgs.un.org/goals>
- the Whitehouse <https://obamawhitehouse.archives.gov/administration/eop/ostp/grand-challenges>
- and CSIRO <https://www.csiro.au/en/about/challenges-missions>.

We suggest that the Roadmap leverage this kind of experience and analysis available on how to successfully implement a challenge, such as <https://www.ida.org/-/media/feature/publications/u/us/use-of-grand-challenges-in-the-federal-government/d10699final.ashx>.

The challenge framework proposed in the Roadmap Exposure Draft does not seem to live up to its potential. The current drafted challenge framework resembles more an industry policy/strategy. The Committee acknowledges industry strategy as an essential and necessary component of a challenge framework but not complete or sufficient in and of itself.

For example the industry sector approach of the current draft encourages siloing of efforts into industry sectors and consequently into science disciplines or public service departments. One effect of a true challenge is to break down silo thinking and marshall resources from various sectors/fields/areas of administration and integrate them into a solution.

Equally, a truly challenge-based approach would better position the Roadmap to achieve its aspirations for impact embodied in Principle 1: "NRI maximises the capability of the research and innovation system to contribute to economic outcomes, national security, social wellbeing and environmental sustainability". A true challenge framework results in a focus on the positive impacts for society. The current industry-based approach is narrower than what is required to achieve Principle 1. A true challenge-based approach would ensure that even industry-focused activities and outputs are actually taken through to societal benefits.

**5. Do the principles articulate the vision and key elements required of NRI, including investment?**

The Committee suggests a slight change to the wording of the principle: "NRI includes the people, skills and knowledge, processes and equipment required to realise the value of the NRI. High quality national collections of data are recognised as infrastructure when they are productive research assets."

Principle 3 states "NRI includes the people, skills and knowledge, data, processes and equipment required to realise the value of the NRI" Whilst the Committee applauds the recognition of the importance of data, this wording falls short. Data itself is infrastructure. It is not something separate from infrastructure "required to realise the value of the NRI". It is the NRI. For example a well curated national longitudinal dataset fits the definition of NRI perfectly as "significant assets … that support leading-edge research and innovation". A broad variety of researchers, private sector scientists, public servants and members of the public use these datasets for a variety of downstream productive activity. Although these data assets do require some maintenance, they are largely an appreciating asset for Australia's digital economy, since with every passing year they provide greater insight into the trends of Australia's physical, natural, social, and built environment. The US National Institute of Standards and Technology defines a data asset as

"Any entity that is comprised of data. For example, a database is a data asset that is comprised of data records. A data asset may be a system or application output file, database, document, or web page. A data asset also includes a service that may be provided to access data from an application. For example, a service that returns individual records from a database would be a data asset. Similarly, a web site that returns data in response to specific queries (e.g., www.weather.com) would be a data asset."

**6. The NRI Roadmap has a clear focus on identifying the NRI investments required to support Australian research over the next 5 to 10 years. Are there any national research infrastructure needs missing in the draft Roadmap?**

An integrated national research repository infrastructure, that caters for both deeper domain specialisation but at the same time can support transdisciplinary research, and enable interoperability of datasets across the physical and social sciences, humanities and medical domains. Any publicly funded repository should have international accreditation/certification by groups such as CoreTrustSeal and the World Data Service. All (meta)data should be compliant with international standards and allows for deeper disciplinary research. We need to invest in both the physical and the social infrastructure that will enable Australian researchers to participate in global data networks designed to address global societal challenges such as the UN sustainable development goals, Climate change and future pandemics.

In Section 4.3 the NRI roadmap has identified a need for an infrastructure for physical samples. We applaud this inclusion but would like to have solid Earth samples specifically mentioned, particularly those collected from expensive drilling campaigns both onshore and offshore to ensure their preservation as national data assets. Where these sample repositories are to be housed is for debate, but currently few institutions can support the infrastructure to curate physical samples and make them accessible for future generations.

We feel that the Exposure Draft does not adequately address support for a modern, open scholarly communications infrastructure. This is urgently needed to support the needs of researchers, funders, policymakers and the Australian public that need to publish, synthesise and access research. All components of the scholarly publishing ecosystem need support, including support for wholly open access academic-led publishing, the long-term development and modernisation of institutional repositories that curate and preserve Australian universities' research, and academic-led experimentation in open research options.

**7. A key priority for Australia is to enhance research translation. The 2021 NRI Roadmap identifies some reforms and investments to achieve this. What other reforms would help deliver this priority?**

Translational research also requires integrated monitoring of social and human impacts of new developments at the time they are released. Including broader stakeholders and beneficiaries early in the planning phases has two advantages: 1) ensuring the needs of eg industry partners are prominent in the design process 2) avoiding the development of new technologies that have detrimental social impacts.

Most universities can only carry out research and analysis of data in small or regional areas. They need access to data at a national scale to refine context and help determine applicability to industry in other areas. Unfortunately in several domains (e.g., Earth Science, Environmental science, health) the large volume, national datasets are controlled by government research agencies and notoriously difficult to access in both minimally processed forms and in modern, non-proprietary, machine-readable formats. We need better integration of Research and Government data assets so as to enable an effective 3-way partnership between to enable greater translation of research breakthroughs to industry.

We need to reform policies and strategies for managing data from the research, industry and government sectors to ensure that they are interoperable between each sector, and comply with best practice internationally. Confidentiality can still be provided to industry, but the publicly funded research data assets from both the research community and government agencies needs to be made openly accessible as soon as it can.

**8. The Roadmap proposes that Australia could make landmark investments to drive step changes in research and innovation over the next 10 to 15 years. Do you agree with the assessment of potential areas for investment in the report? What other areas do you consider might fit the definition of landmark investment?**

Research translation investments at the scale proposed in the step change requires a partnership between Research, Government and Industry. In several disciplines translating research is limited because the research community cannot access vast government data assets in formats that are suitable to modern research, but can also be used in future 2030 computing which is likely to be at exascale.

**9. Please add any other comments you would like to provide to the Expert Working Group.**

The NRI Expert Working Group will need to have wide membership to reflect the spectrum of providers, users, and beneficiaries of NRI. This function should also not detract from the established clear decision making process of the Minister and Department nor with the well established Roadmap and Investment Plan cycle.

**10. If you have a PDF (.pdf) or Word document (.doc or .docx) to share as part of your feedback, you can upload your file here. Please keep documents brief.**

*[Additional document on following page]*

NATIONAL COMMITTEE FOR
**DATA IN SCIENCE**

Australian Academy of Science

## Point 1: Data is Fundamental to Research

Data are fundamental to all domains of research and datasets and data-services are critical research infrastructure[1] in a world where Open Research approaches have proved fundamental in addressing the global challenges we face. **Continued funding of ARDC is thus fundamental to ensuring a "whole of system" integrated response** is maintained in Australia, including all research domains, other sectors (industry, government), and across multiple scales (from exascale to long tail).

## Point 2: "One size" does not fit all

Many fundamental data principles are shared across the Australian research landscape (e.g., FAIR[2], CARE[3]). However, how data are managed, served and used is commonly domain-specific. Interdisciplinary variation must be supported by the infrastructure since no single facility or approach will suit all. Furthermore, because of different resourcing there currently is a wide disparity in the application of these principles across disciplines.

In computationally intensive domains such as bioinformatics, climate, oceanography, solid earth and astronomy, **investment in exascale data capacity situated within NRI high performance computing environments is essential**.

At the same time, coordinated investment in the diverse needs of the small scale, "long tail" communities is vital. For middle scales, investment in Australian domain-specific data assets is a priority. All funded data assets should comply with international certification schemes (e.g., CoreTrustSeal[4], ISO16363[5]).

**Investment in NDRI must cover this spectrum of different needs with a corresponding spectrum of capabilities**.

## Point 3: People are Essential

People are an essential component of data infrastructures: they allow for the creation and curation of data in a form that can facilitate interdisciplinarity and ensure reuse of any prior investments.

[1] The term 'data' covers 'digital objects'— research data, metadata, bespoke algorithms, workflows, models, and software (including code)—resulting from the research - see Recommendation of the OECD Council concerning Access to Research Data from Public Funding
[2] The FAIR Guiding Principles for Scientific Data Management and Stewardship
[3] The CARE Principles for Indigenous Data Governance
[4] CoreTrustSeal Certification for Trustworthy Data Repositories
[5] ISO16363 certification of Trustworthy Digital Repositories

**NCRIS facilities need to be funded to employ experts** who can work with scientists to optimise their use of the infrastructure, including the generation and curation of data to world leading standards. Building data science skills should be integral to any NCRIS infrastructures and linked with strategies that generate/extend National Data Assets and which increase capacity in data science skills across the research sector.

## Point 4: International and National Standards are Critical to Convergent Research

To move beyond domain-focussed research and towards machine to machine data interactions in convergent research, community standards in data are fundamental.

**Every NCRIS facility should therefore be resourced to adopt data and metadata standards** complying with international best practice that support sharing of data and also its integration and linkage to other parts of the research ecosystem such as publications and code. Every facility should **adhere to a data governance and management plan** that enables cross disciplinary data sharing in machine to machine environments. Finally, every facility should, appropriate to their global domain, **adopt common approaches to "informatics" (eg, scientific terminology, data models and metadata) supported by ARDC**, as one of its essential "whole of system" roles within national research infrastructure.

## Point 5: Data Partnerships between Research, Government, and Industry are fundamental

The flow of data between the research, government and private sectors is critically important for society. As data flows among all three sectors, benefits accrue to all participants, and to the community as a whole. National Research Infrastructure should facilitate that three-way data flow, maximise benefits realised from Australian data assets, and maintain public trust.

**Through ARDC, the Roadmap should prioritise investments in appropriate research-public-private "data partnerships**", to build, make accessible and/or use high-value national data assets that integrate data in societal mission areas such as health, human services, demography, education, agriculture, transport, housing, resources (groundwater, minerals, energy) and geo-location as well as national manufacturing priority areas.

**- National Committee for Data in Science, December 2021**