Online submission

1 December 2023

Australian Academy of Science submission to the *Draft NDRI strategy consultation*

National research infrastructure is the linchpin pushing Australian science, especially in the digital era. In every facet of research, from medicine to climate science, the integration of digital tools has become inseparable from innovation and discovery.

The collaborative power of a national infrastructure not only accelerates breakthroughs but also ensures that advancements are accessible across disciplines. With the exponential growth of data and the demand for sophisticated and powerful computational capabilities, a robust digital research infrastructure is a necessity.

It is the accelerator pedal for our scientists, providing the speed and agility required to navigate the complex challenges of the 21st century and allowing Australia to be an active participant in, and beneficiary of, global scientific knowledge.

The Academy recommends that the panel looks for opportunities to be more ambitious than has been put forward in the strategy. Namely the NDRI strategy should:

- Replace the term "user-centric" with "community-centric" to better convey the emphasis on coordination, established best practices, standards, and shared philosophy within research teams.
- Commit to horizon scanning to identify Australia's future NDRI challenges and needs (known and unknown) and provide direction on priorities and investment, leveraging existing expertise in the Learned Academies.
- Recommends that a long-term computing strategy is developed that details the essential role of High-Performance Computing (HPC) to Australian science, to position Australia to establish Tier 0 capability.
- Have a focus on interoperability where imposing standards across scientific disciplines is not practical or possible.

Vision for a future NDRI ecosystem

The vision for the future NDRI strategy underscores the need for a community-centric approach.

The Academy recommends replacing the term "user-centric" with "community-centric" to better convey the emphasis on coordination, established best practices, standards, and a shared philosophy within and between research teams.

The intention is to attract users into an ecosystem of best practices, including standards, while aligning more closely with the active role of research communities.

There is an opportunity to embed the UNESCO Open Science Recommendation into the NDRI vision, and throughout the strategy. This involves not only referencing data and open software but encompassing all elements mentioned in the recommendation. By doing so, the vision can better capture the comprehensive nature of the NDRI ecosystem and its alignment with global open science principles.

The vision should highlight the importance of visibility for the NDRI ecosystem, and emphasise the need to articulate its capabilities, accessibility, and avenues for influence. This concept could be seamlessly integrated

into the vision by incorporating the term "visibility", ensuring that the strategy emphasises the crucial aspect of making the NDRI ecosystem known and accessible.

The vision should articulate the active role of scientific research communities in the development of digital research methods. It should encourage their active involvement in skills development, the initial development of analytic software tools, and the generation of scientific data. By recognising this active participation, the vision can more accurately reflect the dynamic and essential role played by research communities in shaping the NDRI strategy.

Government (public) data is a valuable resource for research and evidence-based policy development. The strategy indicates the requirement for "whole-of-government approach" but there is a lack of information about how the NDRI strategy integrates with other government initiatives such as the DATA Scheme, *Digital Economy Strategy*, the *Australian Cyber Security Strategy* and the *Data and Digital Government Strategy*.

Customarily, the Australian Government has collaborated with Australia's learned academies to undertake horizon scanning via the Linkage-Learned Academies Special Projects, the objective of which is to support the development of Australian research.

The Academy recommends that the strategy commit to horizon scanning to identify Australia's future NDRI challenges and needs (known and unknown) and provide direction on priorities and investment, leveraging existing expertise in the Learned Academies.

Outcome 1 – Underpinned by training frameworks for researchers and NRI

The Academy supports the focus on developing skills and careers for individuals crucial to providing technical expertise in digital research infrastructure.

Skilled workforce and clear career pathways are needs discussed in Academy reports: *Advancing Data Intensive Research* and *Australia's data-enabled research future: Science*. A key challenge identified is the establishment of career paths for early career researchers specialising in NDRI, including areas like high-performance computing, data management, analytics, visualisation, and software engineering.

Addressing this challenge requires not only training researchers and the NDRI workforce but also necessitates a general uplift of digital literacy across disciplines. The example of ACCESS-NRI illustrates the need that will soon become a reality for many disciplines, where highly skilled digital researchers need work in partnership with specialist workforces to manage software and data.

Many examples of thriving areas of NDRI activity are applied science, however the vital role of fundamental science in driving research and development in big data methods, high-performance computing, statistical inference, and machine learning should be recognised in the NDRI strategy. Without these capabilities, Australia will not be able to meet national ambitions in applied or translated research.

Outcome 2 – Responsive to disruptive technological and societal shifts

Australia is currently lacking an ambitious Tier 1 High-Performance Computing (HPC) and High-Performance Data (HPD) strategy.

Our Tier 1 computing facilities should be viewed as more than just computer environments – they are laboratories fostering expertise and facilitating exploration of new technologies.

The draft NDRI strategy mentions HPC only twice, which is inadequate recognition of the vital role of a Tier 1 facility in computing, data-intensive research, and the national capability for implementing AI and machine learning developments to progress areas of Australian science across all disciplines.

The strategy should also look to fully address researchers' needs for competitive access to resources on existing Tier 1 facilities.

The Academy recommends that a focused long-term strategy is developed that details the essential role of HPC to Australian science, to position Australia to build Tier 0 capability.

Outcome 3 – Consistent in its standards for data collection, curation and access

It is unlikely that the proposed single, consistent framework for data standards across all research disciplines would be practical if it is overly prescriptive.

Large-scale international projects, such as the Large Hadron Collider experiments, inherit procedures and hardware specifications from entities like CERN, making it unfeasible for Australian science to establish consistent frameworks across every single discipline. Even within disciplines funded by the NCRIS, establishing a single framework may encounter challenges due to variations in data storage methods, metadata needs and researchers' literacy in processing raw data.

The Academy recommends:

- An emphasis on interoperability where blanket standards are not possible across disciplines, with universal implementation of FAIR and CARE principles across the NDRI system.
- Recognising and providing the research funding to cover the costs of managing data and complying with FAIR and CARE principles.
- An explicit separation between data taken and stored within Australia versus data from overseas, acknowledging the diversity in data storage practices.
- Development of a coherent framework for data curation that is flexible and future-proof.

The UNESCO Open Science Recommendation should be noted for its relevance in setting standards for data collection, curation, and access, particularly addressing the exponentially rising volumes of data generated through modern research practices.

The draft strategy notes that there will be co-locations of data and compute where necessary, and it should be noted that this will increasingly become a fundamental and unavoidable requirement for many areas of science.

Outcome 4 – Integrated across levels of computing and data

In the fields of climate science, weather, and oceanography, the effective utilisation of data, especially in the petascale range, relies heavily on Tier 1 HPC facilities. However, Australia currently lacks an integrated strategy for handling large-scale data, particularly at the petascale and beyond.

The absence of co-location of Tier 1 HPC systems with data storage poses a significant limitation on the scale of research endeavours in these areas. The computational challenges involved underscore the need for increased interoperability, including integrated metadata, to fully harness the potential of such extensive datasets.

In areas of climate science, weather and oceanography, climate projections data is only fully useable when it is within a Tier 1 HPC facility. The data for these areas of research is routinely petascale, and the computational challenge of fully utilising this data requires clear and ubiquitous standards that includes integrated metadata.

The integration of institutional facilities with national facilities may be able to enhance technical support, particularly for universities running HPC systems. This integration, coupled with a national network of expertise may provide benefits, especially considering the typically low staffing levels for technical operation in university based HPC systems.

Outcome 5 – Cybersecure, particularly for national-scale data and computing

Despite many areas of research making their data and code open, Australia's national research infrastructure may be targets for cyber-attacks. The proposed policies are largely sensible and proportionate. The emphasis on cybersecurity, particularly for national-scale data and computing, lacks explicit mention of varying security requirements for different data categories.

The necessity for highly secure data enclaves, especially for the storage and analysis of sensitive human-related data, is not explicitly addressed. National-scale data is not always the highest risk in terms of cybersecurity—the degree of sensitivity, sovereign importance and potential re-identification risk should be equally, if not more, critical than the scale of the data.

Outcome 6 – Maximised by openly available research software tools

The focus should shift from open software tools towards the adoption of the entire UNESCO Open Science Recommendation.

The strategy should position open science as key to NDRI, and emphasise the benefits of wider user bases in improving and developing open-source software. It highlights successful examples in Australian physics, where world-leading open-source codes enhance the visibility of Australian science, suggesting the importance of supporting career paths in this domain.

To discuss or clarify any aspect of this submission, please contact Mr Chris Anderson, Director Science Policy at Chris.Anderson@science.org.au.