Submission to the Draft National Digital Research Infrastructure Strategy

The Academy of the Social Sciences in Australia (the Academy) is an independent, not-for-profit organisation that brings together the multidisciplinary expertise of our nation's leading thinkers to provide practical, evidence-based advice on important social issues facing society.

As the pre-eminent organisation in Australia representing excellence across the social science disciplines, we welcome the opportunity to respond to the Draft National Digital Research Infrastructure Strategy (the draft Strategy).

Overview

The Academy supports the development of the National Digital Research Infrastructure Strategy (NDRI Strategy) and its objective to articulate an explicit set of strategic aims and principles to coordinate national digital research infrastructure (NDRI) investments and decision-making.

This submission presents the Academy's response to the consultation questions.

Our views and recommendations are informed by our work on social science research infrastructure needs and capabilities, including extensive consultation with sector stakeholders on the development of a decadal plan for social science research infrastructure.

With this background, the Academy makes five key recommendations and a number of suggested steps pertaining to specific outcomes and challenges identified in the draft Strategy. Our recommendations are:

- **Recommendation 1**: To include two or more experts on *Indigenous Data Governance and Indigenous Data Sovereignty* on the NRI Advisory Group.

- **Recommendation 2**: To include Australia's five Learned Academies, and other relevant discipline-oriented research community organisations among the key stakeholders driving the design, implementation and review of the NDRI Strategy

- **Recommendation 3:** To set explicit, measurable goals regarding the extent of use of NDRI among the NDRI Strategy's list of critical outcomes

- **Recommendation 4:** To include coordination between NDRI Strategy and Australia's research funding policy (*National Competitive Grants Program – NCGP*) as a priority outcome for realising the future NDRI ecosystem vision

- **Recommendation 5:** That the strategy aspire to cultivate a dynamic and ever-expanding repository of top-tier research data and analytics assets as a priority outcome of the NDRI Strategy.

We would welcome the opportunity to contribute to the finalisation of the NDRI Strategy, and to further collaborate with the Department of Education (the Department) in its implementation and review.

To discuss opportunities for engagement or for further detail on any of the matters raised in this submission please contact Andrea Verdich, Policy Director on 0438 218 352, or andrea.verdich@socialsciences.org.au.

Vision for a future NDRI ecosystem

The Academy agrees that a user-centric design (outlined in the "Vision for future NDRI ecosystem" section) is an appropriate foundation for Australia's NDRI ecosystem over the next 10-15 years.

The following recommendations are made in support of this vision.

> **Recommendation 1.** To include two or more experts on *Indigenous Data Governance and Indigenous Data Sovereignty* on the NRI Advisory Group.

Key points:

- The draft Strategy states: *"Development of the NDRI system must include engagement with Indigenous Australians and acknowledgement of the value Indigenous knowledges can offer to this process"*. A more resolute step towards embedding Indigenous input into Australia's current and future NDRI would see First Nations expertise present at the highest level of NDRI design, planning and decision-making; currently, the NRI Advisory Group.

Suggested steps:

- To include at least two Aboriginal and Torres Strait Islander experts on *Indigenous Data Governance* and *Indigenous Data Sovereignty* into the NRI Advisory Group. Having two or more Indigenous experts as members would ensure different First Nations' perspectives are represented, and would help mitigate against potential negative consequences (e.g., underrepresentation, isolation, disempowerment, superficial engagement, reinforcement of stereotypes, resistance to change, burnout).

> **Recommendation 2.** To include Australia's five Learned Academies, and other relevant discipline-oriented research community organisations among the key stakeholders driving the design, implementation and review of the NDRI Strategy.

Key points:

- Researchers' data and analytics needs are shaped by their affiliation to a specific discipline(s). Consequently, a user-centric strategy should engage researchers not as individuals or members of individual research institutions, but as representatives of disciplinary organisations

- Research disciplines are organised and accessible through various networks and are represented at the national level by Australia's five Learned Academies. These organisations are supported by and relied upon by the Commonwealth to provide independent expert advice spanning research disciplines, and are well suited to playing a role in the design, implementation and review of the NDRI Strategy.

Suggested steps:

- The draft Strategy (*Purpose*, p.5) explicitly names research communities as **users** of the NDRI Strategy but does not include them amongst the stakeholders **driving** its design, implementation and review. We suggest adding:

  "The NDRI Strategy will act as a vehicle to promote and anchor broader discussions across government agencies, state and territory governments, research institutions, ***Learned***

Academies and other research community organisations, public sector organisations and industry. These stakeholders will be essential in driving the national change and high-level outcomes necessary for an optimised and a nationally coordinated NDRI network that is aligned with international developments and best practice".

- On page 6 of the draft Strategy, we suggest adding the following item under "The Government will drive implementation of the NDRI Strategy by...":

"Involving discipline and research community organisations and peak bodies in the scoping of future NDRI capabilities, so that decisions and investments are directly responsive to user needs."

- On page 7 of the draft Strategy, we suggest refining the terms of collaboration with research disciplines and communities by specifying:

"Close collaboration with user communities, through engagement with research community organisations and peak bodies, including Australia's five Learned Academies, is critical to deploying fit-for-purpose solutions and ensuring responsiveness".

Challenges

The Academy notes the statement "*The identified Challenges (outlined under the headings for Outcome 1 to 6) adequately describe the major issues facing the Australian NDRI ecosystem over the next 10-15 years*".

We note the listed challenges and propose that the following are added to the NDRI Strategy. The additional challenges correspond to the additional proposed outcomes (in Recommendations 3, 4 and 5 above).

Proposed outcome	Corresponding challenge
By 2030, Australia's NDRI system will be… *… consistently used by a majority of Australia's research workforce, including users from the majority of research disciplines, and who are highly satisfied with NDRI-driven improvements to research quality, productivity and impact"*	*Australia's NDRI system cannot be solely built for a minority of advanced users, but simultaneously aim to enable the majority of Australian researchers, in the majority of disciplines, to transform the ways they operate in the digital age. In other words, Australia needs NDRI that caters for advanced and less advanced users, is either free or affordable to most users (including HDRs), has capacity to service a large user base, and is geared to achieving the biggest possible quality, productivity and impact improvements for the largest number of researchers.*
By 2030, Australia's NDRI system will be… *… supported by parallel, direct investments in data and software research outputs, through new or expanded research funding streams*	*Australia's current research funding ecosystem does not have funding streams to support the production of such reusable data or software assets and tools. While those funding streams are missing, NDRI is bound to stay underutilised.*

under Australia's National Competitive Grants Program	
By 2030 Australia's NDRI system will be… *… feature a dynamic and ever-expanding repository of top-tier research data and analytics assets*	*Australia's data and analytics producers and custodians face many financial, technical, cultural, regulatory and other barriers to expanding the shared poor of data and analytics resources. While those barriers persist, NDRI is bound to stay underutilised.*

Outcomes

The Academy notes the statement: "*The six outcomes (identified in the outcomes section) adequately capture the priority features of an Australian NDRI ecosystem that will meet the vision (described in the "vision for future NDRI ecosystem" section)".*

We note the listed outcomes, and recommend the following are added for a complete strategy.

> **Recommendation 3.** To set explicit, measurable goals regarding the extent of use of NDRI among the NDRI Strategy's list of critical outcomes.

Key points:

- The user-centric vision outlined in the draft Strategy presently focuses on *design*.

- A user-centric approach starts with design, but is ultimately tested, demonstrated and continuously improved through utilisation and uptake.

- We suggest the addition of explicit aspirations, targets and monitoring/review mechanisms in relation to NDRI use, which:

    o Set clear expectations of NDRI operators, research institutions and users regarding user-centric design and implementation of NDRI

    o Require NDRI operators to measure and report on infrastructure use

    o Provide clear guidelines around the cost and capacity of NDRI use to empower early users (the majority) and to signal to capability providers the expected level of service and acceptable cost-recovery models. Users should not be expected to pay costs for services that have already been funded by government.

    o Finally, we suggest that guidelines and policies should preclude operators of government funded research infrastructure from limiting access based on intellectual property or similar grounds.

Suggested steps:

- To add the following to the list of priority outcomes listed in page 7 of the draft Strategy:

    "By 2030, Australia's NDRI system will be used by a majority of Australia's research workforce, including users from the majority of research disciplines, with an expectation that these users will be highly satisfied with NDRI-driven improvements to research quality, productivity and impact".

Recommendation 4. To include coordination between NDRI Strategy and Australia's research funding policy (*National Competitive Grants Program – NCGP*) as a priority outcome for realising the future NDRI ecosystem vision.

Key points:

- Australia's overall public investment in research projects and personnel (through the ARC, the NHMRC and other National, State and Territory government agencies) and in research infrastructure (through NCRIS, the ARC and institutional investments) is substantial but highly siloed and lacks coordination.

- The draft Strategy proposes investing in NDRI to support the sharing and reusability of data (collection, curation and access) and research software assets. However, Australia's current NCGP does *not* currently fund the production of such assets:

  - Competitive research grants generally do not provide for the additional work required to take data and analytics assets produced within the course of research projects to the standard required for sharing and reuse

  - Competitive research grant schemes do not target the production of nationally-significant data assets, such as national collections, national surveys or longitudinal studies, or supporting tools (research vocabularies, discipline-specific data standards, data-related training, data annotation tools)

  - Competitive research grant schemes do not target the production of nationally-significant analytics assets, such as predictive models (e.g., agent-based model to predict consequences from climate change across multiple domains); large language or multi-format models, or supporting tools (research software standards, related training)

- Proceeding with the Strategy could mean Australia would be investing in capital infrastructure (i.e., through the NDRI Strategy) ahead of implementing the necessary parallel adjustments to its operations (i.e., the funding of research activity through the national competitive grants program). The Academy believes that coordination between research funding and research infrastructure investments is critical to achieving the desired vision and ensure optimal usage and benefits from Australia's NDIR.

- The draft Strategy recognises challenges in training, attracting and retaining an appropriately skilled NRI workforce, including a recommended approach where *"providers and government will work to address staff shortages"*. The availability of research grants to fund research data and software outputs is, possibly, the shortest path to creating meaningful, attractive careers for NRI professionals, and the incentives for research institutions to train and retain them.

- The required parallel changes to research funding policy would involve the creation of new or expanded research grant streams to specifically support the production of research data and software. The production process and production cycles of research data and research software are radically different to those of the traditional research project. Absence of dedicated funding streams for research data and software put a ceiling to Australia's potential to globally lead and innovate in a digital research era.

Suggested steps:

- To add the following outcome to the list of priority outcomes listed in page 7 of the draft Strategy:

"By 2030, Australia's NDRI system will be supported by parallel, direct investments in data and software research outputs, through new or expanded research funding streams under Australia's National Competitive Grants Program"

Recommendation 5. That the strategy aspire to cultivate a dynamic and ever-expanding repository of top-tier research data and analytics assets as a priority outcome of the NDRI Strategy.

Key points:

- The ultimate purpose of NDRI is to support the production and circulation of high value and high quality, reusable research data and analytics assets and tools. Consequently, a critical outcome of the NDRI Strategy should be the cultivation of a strategic, top quality, continuously evolving pool of shared data and analytics assets (produced with support from, and circulated through NDRI)

- Some of the data and analytics assets required for a national pool are already in existence, however historical arrangements and other barriers (cultural, social license, technical, financial, legal or otherwise) prevent their use as national research infrastructure. The removal of such barriers will require a significant investment in education, interoperability standards, data curation and governance arrangements.

- Other data and analytics assets are yet to be produced, and the production of such assets will also require significant investment.

- Consequently, the Academy suggests that the NDRI Strategy should:

    o Explicitly aim for the cultivation and ongoing evolution of a high-quality, high-impact shared research data and analytics asset pool, as a critical outcome

    o Establish fair and transparent mechanisms to allocate investments in data and analytics assets (i.e., mechanisms and processes to guide the sector in identifying the data and analytics assets that best match national or discipline priorities).

    o Directly invest in the acquisition or production of critical data and analytics assets, through dedicated research funding streams (per Recommendation 4).

- Substantial challenges persist in researchers' access to diverse government and commercial data assets essential for domain-specific research. The commercialisation of some government data assets such as Geoscape and Land Registries, further increases these barriers. While NCRIS Facilities like AURIN and AHDAP aim to facilitate access, more work is needed to facilitate timely access to data assets critical to social science research. In order to facilitate better research utilisation of pubic data, it is important that the Strategy deliver mechanisms for whole-of-government research data governance and standard protocols.

Suggested steps:

- To add the following outcome to the list of priority outcomes listed in page 7 of the Draft Strategy:

    "By 2030, Australia's NDRI system will feature a dynamic and ever-expanding repository of top-tier research data and analytics assets."

- To recognise the important role that Commonwealth Government agencies such as the Australian Bureau of Statistics, the National Library of Australia, State and Territory archives, and similar institutions play in providing digital research infrastructure for the social sciences, and to give them appropriate consideration as part of the NDRI Strategy.

Outcome 1. Underpinned by training frameworks for researchers and NRI workforce

The Academy agrees that the content in the *"underpinned by training frameworks for researchers and NRI workforce"* section adequately presents the high-level information expected for this section of the NDRI strategy.

The Academy believes that there is an insufficient emphasis on *analysis* skills in the draft Strategy. Data science, which includes tools like artificial intelligence and machine learning, is based on three foundations – statistical science, computational science and subject matter. The draft Strategy includes multiple references to computational infrastructure, but the role of statistical science is not mentioned.

The Academy suggests that the composition of the NRI Advisory Group be considered in light of the full range of skills outlined in the strategy.

Outcome 2. Responsive to disruptive technological and societal shifts

The Academy agrees that the content in the *"Responsive to disruptive technological and societal shifts"* section adequately presents the high-level information expected for this section of the NDRI Strategy".

The draft Strategy describes elements of responsiveness (e.g., monitoring technological and societal change), but is vague around the mechanisms to ensure such responsiveness. We propose a concrete measure:

The Academy suggests that funding allocations for NDRI enterprises, such as NCRIS should include an end-of-enterprise date, deliverables for the period, and a legacy plan (how assets generated will be preserved or transferred). Obsolescence cannot be avoided. End-of-life-conscious investments provide a mechanism for NDRI to stay responsive.

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

The Academy agrees that the content in the *"Consistent in its standards for data collection, curation and access"* section adequately presents the high-level information expected for this section of the NDRI Strategy. The Academy notes that a sector-wide data management framework will need to handle:

- The privacy, confidentiality and social license issues associated with sensitive human data

- Data retention, including rules to determine length of storage for relevant collections.

The enforcement of any data management provisions included in the framework should occur during the ethics approval process, before data collection commences.

The Academy considers that the framework will remain underutilised if the Strategy does not simultaneously aim for:

- Coordinated, parallel investments in research data and analytics assets, through new, dedicated research funding streams (per Recommendation 4)

- Cultivating a rich, continuously evolving shared pool of high-quality, high-impact research data and analytics assets (per Recommendation 5).

The current description of Outcome 3 includes the need for language identifiers to be used in metadata, specifically the Austlang codes. This is a necessary but insufficient step, and one that precludes consideration of Indigenous languages from other countries. As Australian researchers collect and hold language data from other countries, *the Academy suggests that it is necessary to formulate aspirations (e.g., for metadata) that set Australia's NDRI in the global context (e.g., add language codes from ISO-639-3 and Glottolog, in the case of language data).*

Outcome 4. Integrated across levels of computing and data

The Academy agrees that the content in the *"Integrated across levels of computing and data"* section presents the high-level information expected for this section of the NDRI Strategy.

The Academy reiterates the need for computing capabilities that effectively handle sensitive human data, as fundamental for usability by researchers in the social science disciplines, and

The requirement for "flexible access schemes that enable a variety of users, including periodic and early users" should be backed by concrete provisions regulating what users can be expected to pay to access computing facilities, and the minimum level of access that each individual researcher can expect (incl. for training purposes) (e.g., users should not be expected to pay costs for services that have already been government funded).

The Academy notes that clear guidelines around cost and capacity are set at the national level in order to empower early users (the majority) to set research goals around compute; as well

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

The Academy does not agree that the content in the *"Cybersecure, particularly for national-scale data and computing"* section adequately presents the high-level information expected for this section of the NDRI Strategy.

The draft Strategy contemplates an NDRI-systemwide solution to provide trust and identity capabilities (data access management) across the current NCRIS system. We support the provision of such system, but note it is insufficient. A NDRI data access management solution must seek to make the entire NDRI ecosystem cybersecure, beyond NCRIS facilities.

The Academy suggests that an NDRI data access management solution should aspire, explicitly, to enable secure access to shared data and analytics for a variety of NDRI users, including research institutions, government agencies, non-profit organisations and the broader public/community.

This is critical for the social science disciplines, considering that a substantial portion of data and analytics used in both the social sciences and the humanities are:

- produced by, or co-produced in partnership with these organisations and stakeholders

- produced for the benefit of these organisations and stakeholders, or

- owned by these organisations and stakeholders (data about them, data donated by them).

In addition to data access management, the Academy suggests that an systemwide NDRI solution be developed to safeguard the integrity of national research data assets (protection of data from unintentional errors, data degradation, or malicious tampering).

The costs involved in data integrity preservation are very high for individual organisations to shoulder alone, and expected to increase data assets grow. Such capabilities will need to be similarly accessible to the entire sector (not only NCRIS facilities).

Outcome 6. Maximised by openly available research software tools

The Academy does not agree that the content in the *"Maximised by openly available research software tools"* section adequately presents the high-level information expected for this section of the NDRI Strategy.

Regarding software, it is critical to distinguish between:

(i) Funding for software development (e.g., to create a cross-domain climate model); and

(ii) Funding for infrastructure to support software development (e.g., software enterprise incubator).

The draft Strategy characterises software development as NDRI. In doing so, it flags software development as an activity fundable through NCRIS. Most software development activity is of much smaller scale than could be supported through NCRIS investments, and the Academy is concerned that characterising software development as infrastructure could stifle innovation.

Consequently, the Academy suggests that software development activity for research purposes should be characterised in the draft Strategy as primarily research activity, with a recognition that such activity should be funded largely through new or expanded research funding schemes of existing funding bodies (a parallel measure critical to the success of the NDRI Strategy).

That NCRIS facilities software components is expected and required. However, NDRI funding into NCRIS facilities should not be interpreted as contributing to sector-wide software innovation.