

**Please note:** the substantive content of the 2026 NRI Roadmap Survey begins at Question 20 (with prior questions dealing with administrative and other information).

As such all submissions that are published include the responses submitted from Question 20 onwards only.

Q20.

**Part 2: Research themes**

2.1 NRI comprises the assets, facilities and associated expertise to support leading-edge research and innovation in Australia and is accessible to publicly and privately funded users across Australia and internationally. We are seeking your input on possible directions for future national-level investment - i.e., where the requirements are of such scale and importance that national-level collaboration and coordination are essential.

The [2021 Roadmap](#) used a challenge framework to support NRI planning and investment. With this in mind, consider likely future research trends in the next 5 - 10 years, and with respect to one or more of the 8 challenge areas identified in the 2021 Roadmap as listed below:

- describe emerging research directions and the associated critical research infrastructure requirements that are either not currently available at all, or not at sufficient scale and
- describe current national infrastructure requirements that you anticipate will no longer fit the definition of NRI in 5-10 years.

Do not limit your commentary to NCRIS funded capabilities.

Q21.

**Resources Technology and Critical Minerals Processing**

Q22.

**Food and Beverage**

Q23.  
**Medical Products**

Q24.  
**Defence**

Q25.  
**Recycling and Clean Energy**

Q26.  
**Space**

Q27.  
**Environment and Climate**

Q28.  
**Frontier Technologies and Modern Manufacturing**

Q29.

2.2 The 2024 statement of National Science and Research Priorities (NSRPs) includes outcomes linked to each priority to assist in identifying critical research needed in the next 5 to 10 years.

Consider the priority statements and, with respect to one or more of the 5 priority areas as listed below:

- describe emerging research directions and the associated critical research infrastructure requirements that are either not currently available at all, or
- not at sufficient scale and describe current national infrastructure requirements that you anticipate will no longer fit the definition of NRI in 5-10 years.

Do not limit your commentary to NCRIS funded capabilities, and where relevant, refer to the underpinning outcomes and research identified in the NSRPs document.

Q30.

### Transitioning to a net zero future

Q31.

### Supporting healthy and thriving communities

Maintaining and renewing connections to Indigenous and community languages is crucial for the cultural well-being of millions of Australians, as well as our neighbours in the Pacific South West. A key pathway through which Indigenous and community languages alike thrive is through creative arts and performance, storytelling and literature, and community-led language programs. Research infrastructure that empowers communities to connect with their languages is vital to addressing past injustices, as well as promoting healthy and thriving communities in the years ahead. The Language Data Commons of Australia (LDaCA) is working with communities in Australia and its region to secure culturally significant language datasets to support language maintenance and renewal. However, a key gap at the present time is investment to support the training of local language champions who are essential for mediating between large-scale research infrastructures and the needs of specific language communities in repurposing their language data. Another key gap, relevant to Indigenous communities as well as the multitude of other language communities in Australia, is the lack of linkage between Government (e.g. ABS) and social science infrastructure that surveys and monitors the number of speakers of different languages, and tracks the cultural and social determinates of community health and wellbeing. Given Australia is in one of the most linguistically diverse regions in the world, and the world is rapidly shifting to language-based technologies, there are also huge economic opportunities for Australia in developing world-leading infrastructure for language and text data. Relevant capability could be established in collaboration with relevant NCRIS capabilities (e.g. ARDC, PHRN), building on the infrastructures currently being developed by the Improving Indigenous Research Capabilities (IIRC), the Australian Creative and Historical Futures (ACHF), the Social Sciences Research Infrastructure Network (SSRIN), and Language Data Commons of Australia (LDaCA).

Q32.

### Elevating Aboriginal and Torres Strait Islanders knowledge systems

A central pillar of Aboriginal and Torres Strait Islanders knowledge systems are Indigenous languages. While the Language Data Commons of Australia (LDaCA) is working to secure vulnerable collections of Indigenous languages, and make them more accessible to the communities concerned in culturally appropriate ways, there is a large gap between what can be done with NLP tools (including recent developments in Generative AI and LLMs) for widely spoken languages, and what can be done with Indigenous languages where there are limited numbers of speakers or only limited data available. It is imperative that there be investment in the development of culturally appropriate LLMs and machine learning tools for working with Indigenous languages, as well as support for training Indigenous researchers and communities in effective uses of such tools relevant to their needs. Capability and tools could be established building on the infrastructures currently being developed by the Improving Indigenous Research Capabilities (IIRC) and Language Data Commons of Australia (LDaCA).

Q33.

## Protecting and restoring Australia's environment

Australia continues to face a significant number of environmental and climate-related challenges. In order to effectively meet those challenges we need greater knowledge of long-term changes in our climate as well as our local environmental ecosystems. There are potentially invaluable understandings of Australia's environment and climate changes built into Indigenous knowledge systems, the full value of which remains to be derived. A means of unlocking the potential of those knowledge systems is through the securing and further explication of Indigenous languages which often hold key contextual information. However, research infrastructure for revitalizing and analysing Indigenous languages in conjunction with environmental and climate data in culturally appropriate ways are almost entirely absent from Australia's current NRI system. Historical documents held in archives like Trove also contain vital data about environment and climate in English and other community languages. A capability that addresses the need for culturally and linguistically-informed data linkages could be established in collaboration with relevant NCRIS capabilities (e.g. Atlas of Living Australia, TERN), building on the infrastructures currently being developed by the Improving Indigenous Research Capabilities (IIRC) and Language Data Commons of Australia (LDaCA).

Q34.

## Building a secure and resilient nation

Australia faces an increasingly volatile and uncertain global environment in which mutual security commitments are diminishing. In order to secure Australia's ongoing security it is vital for Government and Australia's defence forces to be able to monitor social and political changes across countries in the Asia Pacific and beyond. This requires, in turn, infrastructure to monitor and assess communications in multiple languages at scale and in real-time. However, there is no such sovereign capability in Australia's current NRI system. There is considerable urgency for such a sovereign capability to be established as Australia cannot rely on or trust privately-owned social media platforms to act in Australia's national interests. A Sovereign capability could be established in collaboration with relevant Government agencies, building on the infrastructures currently being developed by the Australian Internet Observatory (AIO) and Language Data Commons of Australia (LDaCA).

Q35.

2.3 The case for a new NRI capability, or enhancements to existing capabilities, typically emerges through advocacy from research communities clustering around rigorously identified needs and goals. Such a concept could respond to a requirement for novel or expanded capacity within a domain, or across domains, and must be such that it could only be made available with national-level investment.

If you have identified such a requirement, briefly describe the need, the proposed infrastructure capability, the medium-term goals, impacted research communities, and the timeframe over which you advocate its establishment. Your response can include links to relevant existing reports.

Australia is a massively multilingual country, in one of the world's most linguistically diverse regions. More than a quarter of the world's languages are spoken in Australia and its region. However, while Australia's rich linguistic heritage is well documented, much of this data remains inaccessible or under-utilised due to barriers to accessing collections, as well as a lack of access to tools and skills for analysing that data at scale. The goal of the Language Data Commons of Australia (LDaCA) is to work collaboratively with researchers, communities and institutions to develop an integrated national infrastructure for analysing spoken, written, signed and multimodal text collections at scale in order to open up the social and economic possibilities of Australia's rich linguistic and cultural heritage for impactful research with significant benefits to the nation. LDaCA is making available valuable collections of national significance more findable, accessible, interoperable and reusable (FAIR) while adhering to CARE principles; developing the computational infrastructure and tools required to analyse language collections at scale; and increasing the awareness and capabilities of researchers in applying digital methods to language and text data. LDaCA was initiated in 2021 as a national infrastructure project that supports language work and language research through co-investment from the Australian Research Data Commons (ARDC). LDaCA is led by the University of Queensland (UQ) in partnership with AARNet, ANU, Batchelor Institute of Indigenous Tertiary Education, First Languages Australia, Queensland University of Technology, University of Melbourne, University of Sydney, and University of Western Australia. As a key focus is on Indigenous languages, LDaCA adheres to an Indigenous Data Governance framework developed in collaboration with ARDC and IDN. The first two phases of LDaCA (2021-2024: HIR001; 2024-2028: HIR024) have focused on: 1. Developing the social and technical foundations for a national, distributed archival repository for language and text data, including: (a) shared, collaborative data governance and standards framework; (b) shared data access, authentication and authorisation policies, procedures and processes; (c) shared technical infrastructure for curation and storage of language data; and (d) shared technical infrastructure for collection and annotation of language data. 2. Continuing to secure vulnerable and nationally significant collections of Aboriginal and Torres Strait Islander languages, Indigenous languages in Australia's Pacific region, (varieties of) Australian English and migrant languages, and sign languages of Australia and its region. 3. Developing a national data portal for accessing and repurposing language and text data of significance to researchers and communities, both that is held in GLAM institutions, including libraries, archives and museums, as well as language and text collections held in other distributed archival repositories. 4. Establishing an integrated analytics environment for researchers to create fully described, reproducible research on written, spoken, signed and multimodal text in accordance with Open Science principles, and aligned with community expectations for research of practical benefit. 5. Providing training and develop resources for researchers and communities to support best practice in accessing, analysing and archiving language and text data in line with FAIR and CARE principles. Key existing components of the Language Data Commons of Australia (LDaCA) include: the PILARS protocols for sustainable research infrastructure, RO-Crate as an implementation-neutral approach to describing data, and the ONI data portal for making data available to human and machine agents with appropriate security controls for data capture and access, and LDaCA Analytics for language and text analysis, including the Australian Text Analytics Platform (ATAP) and the Language Technology and Data Analysis Lab (LADAL). These components of the LDaCA infrastructure have been purposefully designed to be maximally adaptable to a wide range of research disciplines across HASS and beyond. The goal of LDaCA in the next NCRIS Roadmap is to provide the technical architecture and collaborative blueprint for an integrated HASS NCRIS capability, and to contribute to the technical foundations of an Indigenous NCRIS capability in line with Indigenous Data Governance principles and community expectations. The LDaCA social and technical framework provides for the provision of key cross-cutting services for research with unstructured text data for HASS and Indigenous researchers and communities, including: 1. National Research Data Archival Repository 2. Unstructured Data Transformation and Repurposing 3. AI-Enabled Text Data Capture and Research Workflows 4. Text Analytics: Tools and Workbenches 6. Digital Methods Training and Research Support 7. Data and Infrastructure Governance From 2028-2032, LDaCA anticipates working in collaboration with other key partners to develop an integrated HASS NCRIS capability, as well as contributing to the establishment of an Indigenous NCRIS capability as appropriate. This will involve working strategically with other focus areas of the ARDC-supported HASS and Indigenous Research Data Commons and other relevant existing NCRIS capabilities in order to leverage existing and emerging relationships with key partners, including GLAM (e.g. AIATSIS, NSLA), community organisations (e.g. Language Centres), private industry (e.g. Amazon, Google), international research infrastructures (e.g. CLARIN, DARIAH), as well as relevant Government and NGO stakeholders. This will require a significant new investment that capitalises on investments to date in the development of LDaCA and other focus areas of the HASS and Indigenous RDC.

Q36.

### Part 3: Industry perspectives

**This section is seeking input specifically from industry-based respondents. Other respondents can skip this section.**

Recommendation 6 of the [2021 Roadmap](#) related to improvements in industry engagement with NRI. To complement work on this topic that has occurred since then, we are seeking additional advice on NRI requirements as perceived by current or potential industry-based users.

Q37.

3.1 Have you (or your organisation) interreacted with or used Australia's NRI?

- ☐ Yes
- ☐ No

Q38.

3.2 If so, please briefly outline the NRI capabilities you (or your organisation) have interacted with or used. Do not limit your response to NCRIS capabilities.

*This question was not displayed to the respondent.*

Q39.

3.3 Please indicate your (one or more) primary reasons for interacting with NRI:

*This question was not displayed to the respondent.*

Q40.

3.4 If you answered no, please indicate your (one or more) primary reasons:

*This question was not displayed to the respondent.*

Q41.

## Part 4: Other comments

4.1 Please elaborate on any of your above responses or add any other comments relevant to the development of the 2026 Roadmap. Your response can include reference or links to existing reports that you recommend be considered during the 2026 Roadmap development process.

The establishment of new NCRIS capabilities for HASS and Indigenous researchers and communities will benefit immensely from an integrated, collaborative approach. LDaCA advocates for an integrated, collaborative approach for a number of reasons. First, HASS research data, unlike many other forms of research data, can be repurposed for use by multiple disciplines and communities across both HASS and STEM. Second, there is a serious need for Australian HASS researchers to be able to work at scale if they are to remain internationally competitive. Only an integrated HASS research infrastructure can deliver the scale required. Third, an integrated approach will maximise learnings from prior endeavours to build effective and sustainable research infrastructure. Not every infrastructure project has been successful. We need to leverage those projects that have been successful, as well as learn from those that have been less so. Finally, while applications of digital and AI-enhanced tools and methods are invariably specific to particular disciplines and communities, there are clearly shared needs in capacity building and tools required across HASS and Indigenous researchers and communities. It is not cost effective to replicate the implementation of essentially similar tools and training programs across more than 40 different FoR codes. There have been a number of reviews that attest to the need for an integrated, collaborative approach to research infrastructure for HASS and Indigenous researchers and communities. LDaCA would like to draw particular attention to three of these: • ACOLA (2022). Australia's Data-Enabled Research Future. Synthesis Report. • ACOLA (2022). Australia's Data-Enabled Research Future: Humanities. • ASSA (2024). Decadal Plan for Social Science Research Infrastructure 2024-33.

Q49.

4.2 Optional Document Attachment.

Note: Our strong preference is that answers are provided against the relevant questions in the survey. However, this file upload option is available for submissions in file format, where needed. Please ensure the document includes your name or organisation.

[Language Data Commons of Australia response to NRI survey\\_with links.docx](#)

28.1KB

application/vnd.openxmlformats-officedocument.wordprocessingml.document