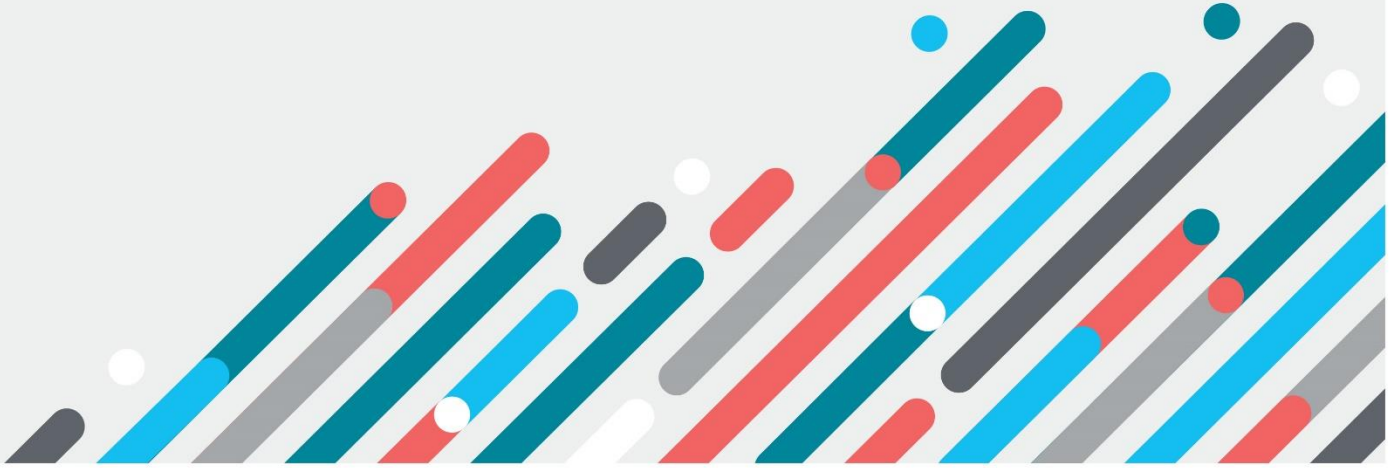
Australian Government
Department of Education

NDRI Investment Plan Consultation Survey Summary

Data



Q1 - What investments could support wide-ranging uptake of FAIR and CARE principles by Australian researchers?	• A national initiative to establish a comprehensive National Research Data Management Framework inclusive of institutions, state and national infrastructure facilities, and publicly funded research agencies, informed by international context and leveraging the ARDC's institutional underpinnings and National Persistent Identifier (PID) Strategy initiatives.
	• Extend, scale up and provide a multi-decadal commitment to underpinning foundational services (national catalogues, PID services, vocabulary and ontology services).
	• ORCID and RAID systems to pull back information, avoiding multiple manual entries.
	• Integration and interoperability between systems (for example, integration of NCRIS facility systems with funders).
	• Incentivising creation of data collections that reduce silos and improve access to linked data.
	• Non-commercial national repository infrastructures that give communities choices about how data will be shared (important for culturally sensitive data).
	• Investing in Resource Description and Access (RDA) as a comprehensive, high-function metadata catalogue consistently indexing data from all repositories (including universities), with intuitive tools for researchers to publish and reuse data.
	• Creating an Australian repository for publishing/versioning software components with rich metadata integrated with workflow tools able to access/use software.
	• Providing data science support for researchers navigating complex research data management landscapes.
	• Supporting career paths for researchers, including for research software engineers, to establish professionals with expertise in FAIR, CARE and Aboriginal and Torres Strait Islander people technology design.
	  ○ High-quality training for researchers to improve their data literacy and data management skills.
	• Digitalisation and aggregation of national humanities and social sciences research assets.
	• Establishing national quality standards for digital research (linked to funding and peer-review).
	  ○ Including accreditation processes for organisations (accounting for research discipline-specific requirements).
	• Funding to support organisations at the point of data acquisition.
	  ○ Consider creating infrastructure (for example, processes, software and hosted repositories with staffing) that forms the link between arrival of data and having it in a FAIR repository.
	  ○ Complying with FAIR principles requires long-term storage capability at the facility where data is generated, particularly when data generated is too large to move.
	  ▪ Investment is required in long-term online (and accessible) storage at national supercomputing facilities. Consider implementing via automatic inclusions for long-term storage with competitive computing awards.
	• Data generated by federal research funding bodies to require it be made publicly available.
	• Simple, universally accessible, and interoperable (with national/local ethics and research governance requirements) data storage and access platform including DOI minting.
	• Establishing metrics to assess the effectiveness of these principles across all disciplines.

	• Investing in next-generation technologies (for example, AI, digital twins, digital biobanking) and align with global standards to accelerate FAIR-compliant research, foster international collaboration, and ensure ethical data practices through CARE principles. • Funding to establish a federated AI capability with local AI nodes to support research environments. • Providing incentives such as awards, recognition programs, or funding bonuses for researchers who publish their data according to FAIR and CARE principles. • Embedding FAIR and CARE fundamentals within university curriculum to target next generation of researchers. • Developing training programs on FAIR and CARE principles that could be adopted at the local level to help promote open data practices. • Investing in awareness campaigns to highlight the importance of these principles through workshops, webinars and online courses. • Establishing collaborative networks to share best practices and resources. By fostering a community of practice, researchers can learn from each other and collectively overcome challenges related to the adoption of these principles. • Supporting investment in free national data hosting services like Nectar for Federal Government funded grantees to overcome cost & infrastructure barriers, especially for large datasets. • Supporting investment in PID implementation at a system level to integrate with trust and identity policy ensures that PIDs are used within the national research infrastructure (NRI) sector. It will inherently incorporate mechanisms to facilitate system-wide adoption of FAIR and CARE policies. • Mandatory reporting on FAIR and CARE compliance with budgetary support for open access. • Using capability maturity models for both FAIR and CARE, including promoting the models and facilitating assessments within Australian research institutions. ○ This would provide common frameworks with which institutions can assess their maturity levels, identify uplift actions, and create implementation roadmaps. • Supporting investment in data annotation tools that guide researchers through the steps of packaging and publishing data assets in alignment with these principles, without requiring any prior knowledge in research data management or FAIR and CARE principles or practices. • Establishment of an Aboriginal and Torres Strait Islander people-led governance body within NCRIS that oversees operationalisation of FAIR and CARE principles. It can develop policies, protocols and implementation strategies for CARE-embedded infrastructure to manage Aboriginal and Torres Strait Islander people data in Australia. • Targeted investments in secure research environment capabilities that integrate the security and data governance standards. • Establishing a national "data library" to register databases availability, and clear information/process how to access.

Q2 - What investments can be considered to address the barriers identified in the NDRI Strategy (pages 16-17)?	• Non-commercial national research repository infrastructures, supported by well-resourced expert teams capable of implementing best practice in relation to data management and governance.
	• Research communities need access to specialist support services and advice.
	• Developing strategies for dealing with very large data sets that are scalable and cost effective.
	• Controlling large language model (LLM) access to data sets, including culturally sensitive data and Aboriginal and Torres Strait Islander people knowledges.
	• Australia needs to invest in technical infrastructure, including personnel, that is tasked with setting standards for digitisation, data management, software/interface design, data analysis, etc.
	  ○ This can be required for peer review funding applications and resulting research.
	• Considerable investments into computational power (for example, supercomputing and machine learning models).
	  ○ Consider investments into quantum computing methods and productivity to aid the above.
	• Funded, dedicated research-data management positions at research organisations, solely for supporting data sharing, management, integration and access.
	• Supporting organisations to move data from ad hoc storage into secure, FAIR repositories that can be federated.
	• Some research fields find that most barriers are not a serious problem. For example, simulation data are written in well-documented, open, public formats with extensive metadata and no access restrictions. However, main barriers are technical, including data size that makes it impractical to send and lack of long-term storage.
	  ○ Invest in long-term storage co-located at national facilities.
	• Establish a range of incentives to open data sharing (for example, regulatory and financial), including funding time required to share data safely.
	• Investments in Fast Healthcare Interoperability Resources (FHIR) will mitigate incompatible metadata and siloed information in local settings.
	  ○ Consider federated systems, with centralised training methodologies deployed to centres of excellence to train on local health data repositories.
	  ○ Investment needed to support resources for use of Patient-Reported Experience Measures (PREMs) and Patient-Reported Outcome Measures (PROMs), including their interpretation.
	• Regarding data generated but never analysed/interpreted, invest in data accessibility with a customisable interface where end-users can design a query to answer a local question or investigate a local problem.
	• Digitising the clinical trial and research management platforms to ensure that data currently housed in paper records within the public health system become a part of the available datasets.
	• Establishing archiving solutions to ensure that data preservation and reuse are available for the long-term.

	• Investments need to be made in supporting projects which bring data together, across sectors, and make them freely available. ○ Such initiatives will improve the accessibility and usability of data and will also create a more integrated and collaborative research environment. • Investment in NDRI processes that efficiently support researchers in the design, planning, and execution of research projects that ensure appropriate management of data, including sensitive clinical imaging data. • Adoption of common metadata standards through guidance and accreditation from bodies like the Office of the National Data Commissioner could help improve interoperability. • Centralised communication strategies and shared data environments could help address information siloes and overload. • Investment in training programmes to ensure that our workforce has best practice data and computing competencies. • Investment in credible career pathways for data, software, and platform specialists to build stable and long-term careers in research infrastructure. • Investments should establish interoperable platforms and trust frameworks to integrate previously siloed systems for seamless data sharing between institutional, national, and international resources. • The NDRI should invest in a comprehensive data management strategy across NRIs to provide clarity on the data management policy including storage, preservation and publication policies. • Develop a centralised, user-friendly data discovery platform that aggregates datasets from across institutions and domains, helping researchers find relevant data more efficiently. • Investments in interoperable data management tools to address barriers in the NDRI Strategy, support FAIR and CARE principles, and enable consistent data management. • Investment in providing stable and PIDs for biodiversity information would increase transparency along the data chain (from publication of a nomenclatural act through to its use in applied outcomes. • Merit allocation schemes giving significant weight to proper publication of research outputs in a form that is truly reusable such as data collections hosted with compute/storage facilities will address poor data management practices. • Investment in making anonymised clinical datasets public and easily accessible to researchers to support scientific and translational research discoveries. • Investment in national participant panel to test survey methodology and validate constructs. • Bringing a not-for-profit and third-sector's data into the research ecosystem for broader consumption and to improve equity.

Q3 - How can future NDRI investments ensure recognition of the importance of data self-determination and the principles of Indigenous Data Sovereignty, in alignment with the Australian Government's Framework for Governance and Indigenous Data?	• Providing researchers with some guarantee that data will not be used to train AI or LLMs without explicit permission and genuine, informed consent.
	• Targeted funding for Aboriginal and Torres Strait Islander research software professionals, development of data standards/protocols and technology design.
	  ○ Aboriginal and Torres Strait Islander people aspects of computational research delivery should be viewed as an asset for Australia, as much as a moral requirement.
	• All data generated by and from individuals should recognise the importance of data self-determination. Funded projects using or hosting data from individuals should adhere to a set of principles include:
	  ○ Clear options and process for withdrawing data or restricting its use.
	  ○ National standardised expiry dates for personal data.
	  ○ Accreditation process for organisations to demonstrate that they are adhering to data sovereignty principles.
	• Must recognise local data sovereignty and importance of including data in research to reduce risk of bias towards other forms of data.
	• Federated solutions are a way to invest in local compute with national methodology to protect data sovereignty.
	• Embedding training and governance (in line with the Australian Government's Framework for Governance of Indigenous Data) at any interface provided that trains users in recognition and awareness of Aboriginal and Torres Strait Islander people data sovereignty and ethical research, including access to 'discovery' data. Must connect with Aboriginal and Torres Strait Islander people to complete this research meaningfully.
	• Developing and implementing data governance frameworks that align with the CARE principles.
	• Resourcing a cross-NRI working group to review the NRI implications of Nagoya Protocol and Digital Sequence Information on access and benefits sharing in Australia.
	• Investments in digitising and providing access to culturally relevant records like languages through platforms, which allows identified individuals to inform access conditions.
	• Metadata records should comply with the Australian Code for the Responsible Conduct of Research, all relevant Federal and State Acts and in accordance with the AIATSIS Code of Ethics.
	• Engaging in community consultation and partnerships with Aboriginal and Torres Strait Islander people communities to foster a respectful and inclusive research environment.
	  ○ Co-design with Aboriginal and Torres Strait Islander communities.
	  ▪ Engaging them in the design, development, and implementation of NDRI infrastructure.
	  ○ Capacity building and training.
	  ▪ Investing in building data skills among Aboriginal and Torres Strait Islander researchers and communities to enhance their ability to manage and use data infrastructure

	• Developing case-studies and exemplar projects to showcase best-practice of self-determination and inclusion would be of benefit to the research community. • The NDRI should support the Productivity Commission's call in the Closing the Gap Review (PC 2024), for a dedicated body to oversee Aboriginal and Torres Strait Islander people data governance. ○ New Zealand's Data Iwi Leaders Group is a case in point.

Q4 - What are the priority data-related NDRI investments that would enhance Australia's research efforts?	• Non-commercial repository infrastructure (including staff) capable of supporting best practice in relation to data governance, metadata and persistent identifiers, rights management and consent for downstream use.
	  ○ Coupled with impartial expert advice and training for researchers charged with navigating complex research data landscapes and opportunities.
	• Digitisation of major data sets.
	• Focus on multi-disciplinary sharing of data to enhance cross-collaboration between fields.
	• Curation of existing and emerging data assets to ensure FAIR compliance.
	• Long-term storage capabilities.
	• Consider following the EU's 'AI Factories' model, where hub-and-spoke approach is being applied to AI across the EU.
	  ○ Investments must be made to make available compute across research centres, including training.
	• To overcome jurisdictional restrictions with sharing data (for example, health data sharing outside of States/Territories), consider local data hubs.
	• Investing in people with skills to analyse and interpret data to link them to and help those who have the questions because they work at the implementation interface.
	• Making data available via digital portals (with filters and dashboards) that allow local customisation of data.
	• Prioritising the development of centralised analytics platforms that can provide real-time insights into research outputs, collaborations and funding trends.
	• Creating online data repositories for historical and cultural records for individual communities so that they can access data about their community and knowledge system.
	  ○ It will enhance Aboriginal and Torres Strait Islander people data sovereignty.
	• Funding of digital biobanks.
	• Integration of large-scale repositories with high performance compute (HPC) infrastructure.
	• Driving culture change in the public service and academia towards enabling public good instead of just risk management when regulating data access and use.
	• Building NDRI capability for data creation teams through investments in training, cross-sector certification schemes, and co-designed university/government programs to upskill professionals in data management, governance, and applying FAIR and CARE principles.
	• Data integration across multiple domains and various levels of government (for example, better integration of state and commonwealth data).
	• Investment in mission critical software and in developing data pipeline capabilities - including scalability and automation that are sustainable and long-term.

	• Developing a national core group of multidisciplinary research software engineers that could build on best practice domain training by providing short-term skills transfer and capacity building, while sharing knowledge from one organisation to the next. • Standardised interfaces and access protocols to enable researchers to move easily between institutional resources and national HPC capabilities • Investing in HPC, interoperable data systems, ethical data management platforms, AI-driven analytics, and advanced visualization tools.