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

Improving the sharing of Experimental Data Metadata Records in Public multi-omics data repository • While this point relates to Translational Medical Research, but is applicable to all of bioinformatics and -omics analyses in general. • A major issue with current publicly available omics data repositories is the quality of the experimental metadata associated with the raw omics data. • Common problems associated with the lack of experimental metadata include: • While public repositories include raw data files, details about the experiments, particularly the experimental groups (e.g., treatment versus control) and covariates (e.g., age, sex, prognosis, phenotype data), required for hypothesis-driven statistical modelling, are not uniformly available in the databases. • If the data are available, which is not always the case, there is no commonly agreed-upon standardized format, nomenclature, or ontology to record the metadata. Even when metadata is available, it is not easily downloadable in bulk or machine-readable, making cross-comparison of data from different databases or large-scale metadata analysis difficult. Often, some of the necessary information is available in the associated published peer-reviewed manuscripts. However, the details linking the experimental groups to the raw data files are frequently buried in supplementary data tables. This prevents automated collection of metadata and large-scale analyses and requires substantial manual re-curation of experimental metadata. In many cases, the metadata and raw data file association information are not available at all, making the data non-reusable and the analysis non-reproducible without further information from the authors. • Our team advocates for the Australian community's involvement in the global discussion on improving the recording of experimental metadata and covariates associated with hypothesis-driven analysis. This includes developing standardized metadata recording formats, nomenclature, and ontologies, which should be uniformly applied to all multi-omics and multi-modal data. • We also advocate for a national repository of multi-omics data to encourage data sharing and reuse within the Australian research community. This will protect the Australian ownership and control of multi-omics datasets and promote collaboration and cross-disciplinary research, including large-scale meta-analyses and the reuse of large cohorts and datasets.

Q24.
Defence

Q25.
Recycling and Clean Energy

Q26.
Space

Q27.
Environment and Climate

Q28.

Frontier Technologies and Modern Manufacturing

High Performance Computing (HPC), GPU, and Australian BioCommons • Currently, Australian BioCommons hosts the Australian BioCommons Leadership Share (ABLeS), which provides researchers focused on biological and bioinformatics research with access to high-performance computing (HPC) resources from the National Computational Infrastructure (NCI) and the Pawsey Supercomputing Research Centre (Pawsey). The access process is simple and streamlined, which is crucial for bioinformatics researchers. Given the rapid technological advances in the multi-omics field and the demand for quick turnaround times, an annual resource request process is neither suitable nor practical. • ABLeS has addressed common bottlenecks that previously hindered researchers, such as the need to provide benchmarking data for software resource requests and the requirement for complicated grant proposals. Instead, ABLeS offers a very simple resource request form, enabling frictionless access to HPC. However, the resources available under this scheme are still limited for large and multi-year projects, which are essential for big data science projects. This includes access to premium GPUs for deep-learning training and AI development and research. • Our bioinformatics team advocate for not only the continual support of the BioCommons and ABLeS scheme but also increased funding to support future bioinformatics research. This would accommodate more premium, fit-for-purpose resources for bioinformaticians. • In addition, Australian BioCommons supports access to HPC for biologists through Galaxy Australia, which offers graphical user interfaces for a diverse portfolio of bioinformatics tools. Galaxy Australia provides HPC and software access for biologists without extensive programming knowledge, enabling them to analyse their data independently. Galaxy Australia is also the primary platform for bioinformatics training for students and biological researchers who want to learn and perform bioinformatics analyses. Expert bioinformaticians also use Galaxy Australia to trial bioinformatics tools for small tests or prototype runs before deploying them on a larger scale using HPC. • All users benefit from not needing the time and expertise required to install research software tools. However, the Galaxy platform is in high-demand and has limited compute resources, restricting its capacity to run even medium-sized projects, such as assembling a plant genome (which we identified from actual usage experience), which is ideal for its target user base. • Currently, a dedicated team maintains existing software tools and installs new ones on Galaxy Australia (e.g. AlphaFold2), facilitating access to new tools for grassroots biology-oriented researchers and students. • Our bioinformatics team advocate for increased funding to upgrade and expand the computer hardware attached to Galaxy Australia and for the management team that maintains and develops the Galaxy ecosystem and software installations.

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

Q32.

Elevating Aboriginal and Torres Strait Islanders knowledge systems

Q33.

Protecting and restoring Australia’s environment

Q34.

Building a secure and resilient nation

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.

Bioinformatics Training and Career Development: • Bioinformatics is a highly sought-after skillset across all biological fields due to the rapid development of technologies that generate data at an increasing pace. Processing, analyzing, and interpreting this data requires many years of specialized training. Training individuals with a broad range of skills needed for diverse multi-omics projects is challenging. Bioinformatics spans multiple disciplines, including biology, computer science, and statistics, but traditional university degrees typically focus on only one of these areas. Specialized training, often at the master's or PhD level, is necessary to achieve the rigorous output required for many research projects. Support for Core Bioinformaticians: • In Australia and globally, there is an imbalance between the limited supply of bioinformaticians and the high demand for their skills. This is exacerbated by the fact that bioinformaticians possess data science skills that are highly sought after by the data science industry, which offers competitive remuneration, well-defined career development pathways, and job security. • Core bioinformaticians often handle a high volume of work while needing to stay updated with new and emerging technologies. They are expected to work with data from new technologies as they become commonplace. Traditional training for new technologies typically occurs during the master's or PhD phase. Without well-developed and accessible training courses, core bioinformaticians may struggle to find structured and guided training to learn and adapt to these new technologies. Structured, publicly available, and open-sourced courses would be more efficient than self-teaching. • There is little recognition or accreditation for self-training, unlike mature careers with well-developed professional development resources, such as Chartered Accountants who have professional training courses and formally recognized certificates. • Bioinformatics as a career is relatively new and lacks a well-defined career pathway. Early-career bioinformaticians often face short-term contracts with little long-term security. Mid-career bioinformaticians, including professional bioinformaticians, may find their career progression plateauing, with little guidance on how to advance. Training for Domain Experts Requiring Bioinformatics Capabilities: • There is a lack of training resources for individuals traditionally trained in areas related to bioinformatics (e.g., biologists, computational scientists, software developers, and statisticians) who need bioinformatics capabilities for their research. • Galaxy Australia, Australian BioCommons, and Bioplatforms Australia provide training, but not in all required fields. Genomics and transcriptomics, with a large group of practitioners in Australia, are better supported for training non-core bioinformaticians. However, support is limited for -omics areas with fewer practitioners (e.g., proteomics, metabolomics). Advocacy for Training Resources: • We advocate for more full-time equivalent (FTE) resources to support bioinformaticians nationwide in areas with fewer practitioners, such as proteomics and metabolomics. Efforts should support the development of individuals involved in multi-omics, cross-disciplinary, and multi-modal projects. • We advocate for working groups to help define career development pathways for bioinformaticians at all levels, particularly mid-career bioinformaticians, where career pathways are not well-defined. • We advocate for resources to develop bioinformatics training and foster a "community of practice" with the mission of training and up-skilling core bioinformaticians, researchers, and practitioners who wish to gain bioinformatics capabilities. Training courses should be developed, accredited, or peer-reviewed, and the up-skilling undertaken by participants should be recognized nationally and preferably internationally. Emerging Technologies: • The field is increasingly moving towards deep learning and large language models (LLMs). It is crucial to understand how these emerging technologies can be utilized for multi-omics, multi-modal, and cross-disciplinary research. These AI tools should be accessible and open source. o We advocate for substantial investment from the government and co-investment from industry to create a critical mass of inter-linked and systems-level data for training AI, machine learning, and large language models (LLMs). These datasets should be open source, open access, and adhere to the FAIR data principles—Findable, Accessible, Interoperable, and Reusable. Additionally, all raw data, associated experimental metadata, and cohort clinical metadata should be made available. o There needs to be continuity planning around the phased and continual analysis of this dataset of national importance. Potential Applications of Emerging AI Technologies: • Reducing Drug Development Time and Resources: AI technologies can streamline the drug development process, reducing the time and resources required. • Integrating Multi-omics and Multi-modal Datasets: AI can help integrate diverse datasets to model cells, organisms, or large clinical cohorts. This integration can lead to accurate predictions, potentially bypassing the need for animal research and testing, and guiding large-cohort studies with precise predictions. Privacy and Ethics: • There are important questions about privacy, ethics and safety of using AI and LLMs. Advocacy for LLM Development: • We advocate for the development of an LLM specifically aimed at augmenting multi-omics, multi-modal, and cross-disciplinary research projects. This LLM would ideally provide best practices for various types of -omics analyses, covering areas such as medicine and health informatics, agricultural biology, and environmental biology. Ideally, these LLM should be open-source and open access, with seamless usage by researchers.

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.

Proteomics, genomics, and metabolomics capabilities.

Q39.

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

- [x] For expertise or advice
- [ ] Access to research resources or products
- [ ] Access to equipment for research
- [x] Access to equipment for operational reasons
- [ ] Help in translating research
- [x] Access to data
- [ ] Support for clinical trials
- [ ] Other (please specify)

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.

Multi-omics Research Software Development and Maintenance • Open-source Software Development: • There has been significant open-source software development in genomics, transcriptomics, and metagenomics. However, proteomics is dominated by commercial software, which has limitations for research purposes, including: • Most commercial software is built for Windows OS, while most HPC systems use Linux OS. Additionally, these software applications are often designed for desktop versions of Windows, not Windows Server. • The software is not designed for deployment on HPC or compatibility with computational pipelines or workflow languages like NextFlow. This makes it difficult for researchers to run pipelines on HPC or share their entire pipeline reproducibly via workflow languages. There are also concerns about the scalability of Windows-based and GUI-based commercial software. • Commercial software often requires expensive licenses, which may not be accessible to all research groups. • There is often no open-source or free alternative to commercial software. This is partly because raw data is recorded in proprietary formats that can only be read by proprietary software. • While data format converters exist, the integrity of the data is not guaranteed, and changes in data format could lead to significant differences in output results (anecdotally reported). • Even if the software is open-source, there are often issues with parallelizing the algorithms (e.g., bottleneck steps requiring a lot of RAM and specialized hardware). The efficiency and fit-for-purpose implementation of research software crucial for large-scale analyses. • Maintenance of Open-source Research Software: • Maintaining and improving open-source research software requires significant human resources. Often, these applications are developed by higher-degree research students or postdocs on short-term contracts, leading to a lack of continuity in long-term maintenance. • Maintaining research software applications requires experience and expertise from the original developer or research lab, which is not always possible or economical for the open-source community. • Advocacy for Resources: • We advocate for resources to be allocated towards the development and maintenance of research software in areas with limited options (e.g., proteomics). The goal is to provide seamless, truly open-source, and open-access research software for the community. • Efforts should focus on developing open-source software that is widely compatible with HPC and workflow languages like NextFlow. These efforts should be prioritized based on the needs of relevant research groups or communities (e.g., guided by the computational proteomics society via community-wide consultations). • We also advocate for resources dedicated to maintaining critical research software. These resources should be prioritized based on fair representation of importance and impact (e.g., community-wide consult or grant applications.

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.