

2016 National Research Infrastructure Roadmap Capability Issues Paper

Name	Simon Eastal
Title/role	Director
Organisation	National Centre for Indigenous Genomics, The Australian National University

Questions

Question 1: Are there other capability areas that should be considered?

There are important areas of health research that would not accurately be characterised as ‘science’. “Health and medical research” would be a more accurate representation of the first capability than “health and medical science”.

Question 2: Are these governance characteristics appropriate and are there other factors that should be considered for optimal governance for national research infrastructure.

Level of interoperability across the research infrastructure system

Equally important is interoperability with other elements of the R&D ecosystem and with international organisations and initiatives. ‘Interoperability’ suggests transactional relationships. ‘Integration’ may better convey the more substantial relationships that are often needed to work across the system.

Strategic approach to whole-of-life costs including defunding or decommissioning

Identifying the need for strategic approach based on investment (as distinct from a cost-based accounting) might better capture what is required.

Intellectual property and moral rights

In addition to IP management, future national research infrastructure, particularly (but not exclusively) big data in health and medicine, will require new ethical, legal and social models that will be integral to success and therefore should be included within the scope of national research infrastructure.

Question 4: What are the conditions or scenarios where access to international facilities should be prioritised over developing national facilities?

Successful access to international facilities and development of national facilities are likely to be interdependent and mutually re-enforcing rather than competing alternatives. Essential international engagement may depend on local adoption of international technical, quality, regulatory and other standards. National comparative advantage (in specific areas of technology, fields of application,

outcomes, etc.) may be required to attain and preserve a position in the international community of respected peer (rather than supplicant or vassal).

Question 6: How can national research infrastructure assist in training and skills development?

It provides a framework to guide thinking in educational institutions about new requirements for training and skills development. Seismic curriculum changes are likely in many areas in coming years. Close engagement between educational and infrastructure developments will have enormous mutual benefits.

Question 7: What responsibility should research institutions have in supporting the development of infrastructure ready researchers and technical specialists?

This is an essential responsibility. Educational success will depend on effective engagement between educators/trainers and emerging infrastructure initiatives.

Educational and training requirements for the 'big data to knowledge' (BD2K¹) paradigm that is emerging in most areas of human endeavour may not be well understood in some domains. New content and course structures will be need to reflect the shift from a rationalist to an empiricist approach to research and from inferential to predictive data modelling.

Current practitioners in many areas may be unfamiliar with this very different approach to knowledge creation. It will, however, be underpinned by a new kind of research infrastructure, which can, therefore, provide a crucial, if indirect, role in developing Australia's future intellectual capital. It can showcase tangible outcomes and benefits in specific areas obtained through domain-specific application of high-level skills (soft infrastructure), in, e.g., data exploration, representation, transformation, computation, modelling, visualisation and stewardship.

Question 8: What principles should be applied for access to national research infrastructure, and are there situations when these should not apply?

These are varied, but recognising natural monopolies, where they occur, and resourcing them accordingly is important. Many facilities will provide better value if they are not run to maximise profit or in unsustainable attempts at 'cost recovery', which can discourage the collaborative engagement needed for the emergence of communities of best practice.

Question 9: What should the criteria and funding arrangements for defunding or decommissioning look like?

A managed fund with a modified capital budgeting strategy that identifies and accounts for both financial and non-financial returns would provide a rational framework for investment decisions, co-investment strategies and asset management.

¹ See, for example, <https://datascience.nih.gov/bd2k>

Question 11: When should capabilities be expected to address standard and accreditation requirements?

One approach would be to establish world's best practice as a default, recognising that lower standards are sufficient in some areas, but would need to be justified to qualify for investment.

Question 12: Are there international or global models that represent best practice for national research infrastructure that could be considered?

The following organisations are a sample of those relevant to the National Centre for Indigenous Genomics.: Genomic Alliance for Genomics and Health²; The Human Heredity and Health in Africa Initiative³; Genomics England⁴; The Broad Institute⁵; Genome Asia 100K⁶; GenomeQuébec⁷; 23andMe⁸; DeCODE genetics⁹; Meilahti Integrated Biobank Infrastructure (Finland)¹⁰; various infrastructure initiatives of Her Majesty's Government in the United Kingdom of Great Britain and Northern Ireland.

The list is far from exhaustive, and not all those on it are dedicated infrastructure providers, but they have all build exceptional infrastructure that could usefully inform the development of infrastructure in Australia.

Health and Medical Sciences

Question 15: Are the identified emerging directions and research infrastructure capabilities for Health and Medical Sciences right? Are there any missing or additional needed?

An important emerging development in medical research, which is not identified in the issues paper, is direct participant engagement in research. This development—broadly 'participant-centric research'¹¹ has great potential and is an important element in the establishment of repositories of samples and data for multiple, long-term research use.

New infrastructure platforms are required to enable increased participant engagement in the research enterprise. These include new ethical, legal and other forms of soft infrastructure that should be developed as essential components of national infrastructure in this domain.

² <https://genomicsandhealth.org/>

³ <http://h3africa.org/>

⁴ <https://www.genomicsengland.co.uk/>

⁵ <https://www.broadinstitute.org/>

⁶ <http://www.genomeasia100k.com/>

⁷ <http://www.genomequebec.com/>

⁸ <https://www.23andme.com/en-int/>

⁹ <http://www.decode.com/>

¹⁰ <https://www.fimm.fi/en/services/biobanking-facilities/mibi>

¹¹ Kaye J, et al. (2012) From patients to partners: participant-centric initiatives in biomedical research. *Nature Reviews Genetics* 13: 371–376.

Question 16: Are there any international research infrastructure collaborations or emerging projects that Australia should engage in over the next ten years and beyond?

See the list provided in response to Question 12. Perhaps a more important question is what general infrastructure (including soft infrastructure) elements are required to ensure Australia's capacity to engage in important new international initiatives.

Question 17: Is there anything else that needs to be included or considered in the 2016 Roadmap for the Health and Medical Sciences capability area?

There is a special need for data stewardship (see Question 35) in health and medical research because of the complexity of data, the sensitivities that arise from its connection to patients and participants, and the complicated political, jurisdictional, organisational and regulatory environment in which it exists. Consideration should be given to the creation of a national body that can direct/coordinate standardisation of data quality assurance, metadata, and more general aspects data management.

Environment and Natural Resource Management

Question 20: Is there anything else that needs to be included or considered in the 2016 Roadmap for the Environment and Natural Resource Management capability area?

The National Centre for Indigenous Genomics is creating, as part of the national research infrastructure, an Indigenous governance framework that ensures inclusion of Indigenous Australians in the conduct and benefits of human genome research. A similar framework should be built into the national research infrastructure to ensure Indigenous inclusion in the conduct and benefits of research in environmental and natural resource management fields.

Underpinning Research Infrastructure

Question 30: Are the identified emerging directions and research infrastructure capabilities for Underpinning Research Infrastructure right? Are there any missing or additional needed?

The strong focus on digitization may cause the full complexity of creating useable digital archives to be under-appreciated. Resources invested in digitization without commensurate investment in related data and knowledge management technologies will return little value.

The excellent brief summary in the bottom right cell in the table on page 48 should be expanded, with digitization recognised as just one component of the process. The importance of technologies that provide trust, reliability, meaningfulness, context, persistence, community/public engagement, and linkage to physical reference collections and archives should be expanded.

Question 32: Is there anything else that needs to be included or considered in the 2016 Roadmap for the Underpinning Research Infrastructure capability area?

With the possible exception of Section 8 (Understanding Cultures and Communities), the importance of digital and linked physical reference collections is understated and needs to be emphasised more strongly both generally and in specific sections.

Nationally supported systems of resource discovery would be particularly valuable in Australia's geographically dispersed research environment. [Eagle-i](#) is a model of such a system operating in the biomedical field in the US.

Data for Research and Discoverability

Question 33: Are the identified emerging directions and research infrastructure capabilities for Data for Research and Discoverability right? Are there any missing or additional needed?

Building on existing capabilities is desirable and necessary. However, the architecture of data infrastructure required in different areas varies enormously. Requirements in health and medicine, for instance, are quite different from those in physics and chemistry, or in social sciences and humanities. It is important that these differences are recognised and accounted for. A 'one-shoe-fits-all' approach would result in sub-optimal, even unworkable, outcomes in some areas. Explicit modelling of the requirements of different research domains to guide resource deployment would improve outcomes.

Question 35: Is there anything else that needs to be included or considered in the 2016 Roadmap for the Data for Research and Discoverability capability area?

For biological data derived from humans, the relationship between donors, donated samples and data derived from samples is changing from transaction to engagement – it is becoming longer-term, more reciprocal and more dynamic. There is a need for infrastructure that supports these new relationships, for example, through research platforms that include participant portals and dynamic consent capabilities.

The dynamic nature of data infrastructure creates a requirement for data stewards, whose role includes management of metadata; data elements, structures, attributes; meta-information about provenance, ownership & related rights, standards, controls; registration, documentation etc. The national research infrastructure should include provision for the training and deployment of data stewards, who extend the function that librarians have traditionally performed—infrastructure that has been the bedrock of intellectual progress for thousands of years.

Other comments

1. An Indigenous perspective may be more important and relevant in some areas than in others. It should, however, be applied to all areas of national collaborative research infrastructure policy and not confined to specific area such as Understanding Cultures and Communities. An Indigenous voice should inform how we, as a nation, support (through national research infrastructure) the creation and use of new knowledge about ourselves, our country and our relationships with the rest of the world and beyond.
2. In the 'Data for Research and Discoverability' area, there is great potential to improve the conduct and quality of research by increasing reproducibility and decreasing the currently widespread dissemination of (very expensive) false research findings. Although not the primary function of infrastructure providers, they are involved by default, and there is an opportunity to make a valuable contribution through requirements for quality and other standards and by providing a focus for communities of best practice.
3. Deployment of blockchain (distributed trusted ledger) technology has great potential to improve: interoperability across the system, collaboration and networking, management of intellectual property, resource management and accountability. It could be particularly useful in Australia's geographically dispersed research system and may provide a valuable national comparative advantage.
4. In many fields the effectiveness of programs to deliver excellent research outcomes and to create value beyond the research enterprise requires integration across many infrastructure components. Governance arrangements must facilitate and must not impede this integration. Australia's research ecosystem consists of organisations with a diversity of legal structures, corporate practices and strategic objectives that are not optimally aligned to achieve high levels of integration.

Robust strategic analysis of the governance, business and management arrangements that are required to integrate national research infrastructure across existing and potential new organisations should be given the highest priority. 'Use-cases' and pilot projects in specific domains that require broad integration across the system would provide the focus needed to identify strategies for effective integration, optimal use of new and existing resources and the delivery of maximum national benefit.