Submission
2016 National Research Infrastructure Roadmap Capability Issues Paper

Name	Richard Neville
Title/role	Mitchell Librarian and Director Education & Scholarship
Organisation	State Library of New South Wales

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

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

GLAM sector infrastructure operates differently to tertiary-based research projects. GLAM clients and researchers, who are often not academics, expect permanent access to open datasets and collections of digitised content. This content is owned and managed by public institutions, rather than generated, as it typically the case for STEM data, by universities and researchers. Currently GLAM data tends to be siloed in each institution's own systems. The extraordinary and exciting opportunity of a robust national research infrastructure would be the capacity to aggregate and search across each of these datasets simultaneously, and to be embedded in national content. This has the potential to be transformational.

While GLAM content may be the subject of competitive ARC bids, in of itself, it is not project-based or likely to become redundant. The State Library of NSW expects that its digital content (in particular that it generates itself from its own collections, rather than commercially purchased on-line resources) will remain permanently accessible to all clients. While GLAM data is less likely to the be size of STEM data, it is generally much more open to public access and would need permanent access to research infrastructure. This suggests that a fundamental core of the research infrastructures needs to be also permanently accessible.

Question 3: Should national research infrastructure investment assist with access to international facilities?

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

Question 5: Should research workforce skills be considered a research infrastructure issue?

Yes. Administrative/professional staff using systems to support data capture, curation, management, preservation and discovery require

specialised skill sets. Specifically, technical skills in implementing and managing infrastructure around data catalogues and repositories. Additionally from a curation perspective, domain specific knowledge of subject matter is required to facilitate effective management of research data.

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

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

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

GLAM institutions, and libraries in particular, have a long tradition of open access to collections, be they original or now increasingly digital surrogates, and born digital collections. There are few opportunities to monetise heritage data, apart from very narrow interests like family history data (e.g. Ancestry.com) or curated datasets (e.g. Adam Mathews or Proquest). Income which libraries or archives may derive from the licensing of this data is small. The GLAM sector expects its collections to be widely accessible, and while it is happy to contribute its content to research projects, it would not want to enter into agreements built around exclusivity of access to collections.

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

Question 10: What financing models should the Government consider to support investment in national research infrastructure?

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

Standards and accreditation have less relevance to humanities research projects. The audiences for this Library's content is much broader than the tertiary sector. Our audience will include academics but it will also embrace independent researchers, creative producers, heritage consultants, community experts and so on, for whom accreditation requirements are not necessary.

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

Question 13: In considering whole of life investment including decommissioning or defunding for national research infrastructure are there examples domestic or international that should be examined?

Question 14: Are there alternative financing options, including international models that the Government could consider to support investment in national research infrastructure?

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?

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

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?

Environment and Natural Resource Management

Question 18: Are the identified emerging directions and research infrastructure capabilities for Environment and Natural Resource Management right? Are there any missing or additional needed?

Question 19: Are there any international research infrastructure collaborations or emerging projects that Australia should engage in over the next ten years and beyond?

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?

Advanced Physics, Chemistry, Mathematics and Materials

Question 21: Are the identified emerging directions and research infrastructure capabilities for Advanced Physics, Chemistry, Mathematics and Materials right? Are there any missing or additional needed?

Question 22: Are there any international research infrastructure collaborations or emerging projects that Australia should engage in over the next ten years and beyond?

Question 23: Is there anything else that needs to be included or considered in the 2016 Roadmap for the Advanced Physics, Chemistry, Mathematics and Materials capability area?

Understanding Cultures and Communities

Question 24: Are the identified emerging directions and research infrastructure capabilities for Understanding Cultures and Communities right? Are there any missing or additional needed?

It is worth noting that Trove offers really important advantages to contemporary humanities research, but is in its current state not scalable or funded appropriately to be developed further. To be a truly national project its capacity would need to be significantly enhanced to be able to ingest or search across the full holdings – not only the bibliographic metadata but also full text OCRed collections – of national collections.

How these national collections can engage with datasets and resources in tertiary institutions is an interesting challenge. In libraries, for instance, there is little genuine (although the situation is changing) collaboration between tertiary libraries and national and state libraries.

There are some cultural differences between collecting institutions and tertiary collections. Libraries, like the State Library, see collection additions to their heritage collections as permanent accessions. Tertiary research is often governed by Ethics committee regulations which can limit, or even deny, access or acceptance for permanent retention.

The Library would strongly endorse comments at 8.3.2 but note that a community managed research infrastructure should not be subject to funding pressures, or conceived in a project framework with the expectation that funding will terminate. Communities will rightly expect permanent access to their data, in a framework which is transparent, stable and clear.

Question 25: Are there any international research infrastructure collaborations or emerging projects that Australia should engage in over the next ten years and beyond?

Question 26: Is there anything else that needs to be included or considered in the 2016 Roadmap for the Understanding Cultures and Communities capability area?

How users can access data – its discoverability – is a key issue for community access. Clients expect easy search solutions. While complex data is often not inherently easy to search or work with, how end users will access and manipulate collection data should always be at the forefront of design. Community users, in particular, do not have the sophistication or tools of experienced tertiary researchers, so systems designed for their use need to reflect this.

National Security

Question 27: Are the identified emerging directions and research infrastructure capabilities for National Security right? Are there any missing or additional needed?

Question 28: Are there any international research infrastructure collaborations or emerging projects that Australia should engage in over the next ten years and beyond?

Question 29: Is there anything else that needs to be included or considered in the 2016 Roadmap for the National Security capability area?

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?

Digitisation per se is not building collections, as suggested in 10.3.6, but rather facilitates access to existing analogue resources and to born digital resources. Digitisation provides previously unimaginable opportunities to interrogate and investigate national collections. But while the emphasis here is on national or internationally significant projects, the value to state or regional spaces needs to be acknowledged. The social impact of projects can be profound at a regional level: the application of a criteria of national significance should be used wisely when assessing the suitability of projects.

There is already a significant national digitisation capacity, but it is true that it is probably not well connected in terms of knowledge sharing. However individual sectors do share and exchange information, which reflects the specific needs of each sector.

Much of the data discussed in this paper will probably be born digital (Word documents, Excel spreadsheets, emails, databases etc), particularly over the next decade or so, and here the capacity issues will be around digital preservation, standards, metadata cross-walks and scalability.

Who has ultimate responsibility for the data output of many of these research projects – particularly the research data – is an interesting question. Are tertiary libraries prepared to accept a role in the permanent retention of some very valuable research data, or will some responsibility fall to collection institutions? Do complex ethics responsibilities for tertiary research data mitigate against its permanent retention?

Question 31: Are there any international research infrastructure collaborations or emerging projects that Australia should engage in over the next ten years and beyond?

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

The chart in 10.3.6 does not sufficiently stress humanities research projects. It would be useful to see the importance of a national research infrastructure emphasise its capacity to engage with the humanities and social sciences: most examples do tend to be STEM focused.

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?

Collaboration to build mature sector based taxonomies and ontologies will facilitate the linking of open data yielding accelerated research and discoverability. In the area of Libraries, there is an opportunity to lead the development of digital humanity taxonomies and ontologies in our region and for Australiana domains.

There is a shortage of data scientists, data engineers, data visualisation and data communicators. These resource capabilities, whilst not an infrastructure component, is necessary in the story telling aspect of digital

humanities. They create digital experiences that are informative, insightful, and interactive.

Question 34: Are there any international research infrastructure collaborations or emerging projects that Australia should engage in over the next ten years and beyond?

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?

Other comments

If you believe that there are issues not addressed in this Issues Paper or the associated questions, please provide your comments under this heading noting the overall 20 page limit of submissions.