CSIRO Social Science and User Experience Working Group Response to The Decadal Plan for Social Science Research Infrastructure 2023-32

To The Academy Of Social Sciences in Australia

Thank you for this opportunity to provide feedback on the Academy's Decadal Plan for Social Science Research Infrastructure 2023–32. The plan reflects an enormous amount of critical, innovative, and well researched thinking that provides a solid foundation for the development of social science research infrastructure.

Please find our comments to specific questions below.

Producing, discovering and accessing data

Q1. How would you modify or augment our description of the current state of assets, systems, rules and skills and training?

- We would encourage reflection on how the plan as a whole captures the dynamism and potential movement of the ecosystem. The work of infrastructure is never completed; it is a constantly ongoing process. The external environment is also constantly in flux (for instance, consider the pace of technical change in AI over the last 10 years). Given that flux is constant, the plan should look to incorporate a set of feedback loops for growth, expansion and improvement that can help it adapt and respond to these changes. This might come in the form of an official Monitoring, Evaluation, and Learning framework, or close reflections on a set of underlying principles that help the plan respond to changes in the material elements identified (assets, systems, rules, skills and training). Constant reflection and review will allow for a more strategic plan that is responsive and resilient to change, and will have greater impact overall.

Q3. Which needs can be met through improvements to existing assets, systems, rules or skills and training? Briefly describe the improvements required.

- While the production, discovery, and accessing of data section notes the importance of community standards, technical regulation, data sharing, and other factors essential for interoperability, how these elements of the data supply chain might be governed is unclear. Thus, we would suggest more consideration needs to be given to the governance of the infrastructure ecosystem, and what this socio-technical system might look like in practice. There is an opportunity to create a governance model as a part of the decadal plan that can help advise a strategic approach, and provide operational advice to determine the specifics of standards and other instruments that are required for

interoperability and other vital functions. This might be represented by building a unique organisational layer into the infrastructure that covers assets, systems, rules and skills, furthering their strategic impact.

Q4. Which needs require that the sector advocates for new assets, systems, rules or training? Briefly describe any new infrastructures you think are required, including where possible examples and any requirements for successful implementation (e.g., incentives, funding, partnerships).

- We note that the current infrastructure ecosystem does not feature many (if any) private sector actors who hold data of relevance to the social sciences (such as socio-demographic data, insurance data, and other forms of personally and socially relevant data). This includes data brokers such as Acxiom or Equifax, or private actors such as Alphabet or Meta. These entities have enormous social science data holdings, and in the case of Alphabet and Meta conduct their own detailed research. There should be a consideration of how these sources are integrated into the social science research ecosystem in a manner that aligns with ethical principles and public good expectations. What levers would be required to achieve this integration remains unclear. Partnerships and other collaborative models might lead to clashes between public good and private profit, and incentives and financial levers would likely be expensive. However, the fact is that private ecosystems exist and are growing, especially as data collection systems like Facial Recognition Technologies and other systems grow in private environments. How these sources are brought into the research environment for public good is an open question and a potential area of growth for social science research in Australia.

Analysing data to generate new knowledge

Q5. How would you modify or augment our description of the current state of assets, systems, rules and skills and training?

- The current infrastructure ecosystem for "analysing data to generate new knowledge" does not seem to acknowledge the different kinds of precarious, invisible, and potentially exploitative systems of knowledge generation that are currently existing in the space. This includes the widespread use of Mechanical Turk and other microwork platforms used to label data, and the outsourcing of labelling work done by private sector actors that may then also filter into the data supply chain. This under-acknowledgement presents ethical issues that the current plan does not seem to engage with. The regulation of AI is, of course critical, but it is the regulation and management of the broader data supply chain that is also required.
- Regarding skills for knowledge generation, the current decadal plan misses an opportunity to consider the role of social science related or adjacent professionals who use social science methods and data to create value and innovation but sit outside of traditional spheres. This includes Strategic Designers, User Experience Designers and Researchers, Interaction Designers, and other professions that are not typically considered or associated as social scientists. The decadal plan might do well to expand its scope to consider these and other emergent professions, given the contributions they make to the economy and society. This includes considering what skills can be drawn from these disciplines to further the impact of social science research and data assets, and how social science research

infrastructure might be made available to these professionals to create potential new innovations and values.

Q6. Can you provide specific examples of data-related challenges your research team faces, where shared infrastructure could significantly boost productivity or support your research aspirations?

- We have a great opportunity through our work to use emergent AI innovations to supercharge traditional qualitative methods (thematic analyses, discourse analyses etc) by using AI and cloud computing capacities to rapidly transcribe or parse text (although technical hurdles remain on parsing large amounts of text). However, we are prevented from achieving this because of privacy requirements, and the fact that most private service provides do not meet the necessarily Australian standards for private data (or our organisational requirements).

Q7. Which needs can be met through improvements to existing assets, systems, rules or training? Briefly describe the improvements required.

- Rigorous privacy, data security, and general data regulation are essential, and should be the focus of enhancement not diminishment. The opportunity is to provide whole of government and/or whole of academia service offerings for AWS/Azure/other cloud capabilities, that comply with Australian data regulations (such as the Privacy Act 1988 and other legislation) so that we can easily access (with support) these services to conduct our research. This support should include assistance to collaborate with legal and regulatory professionals to ensure research practices and applications are legally compliant, while also being fit-for-purpose for research. We have observed that collaborations between legal professionals and researchers are sometimes challenging, with different languages and approaches hampering the establishment of the common ground necessary for collaboration.

Q8. Which needs require that the sector advocates for new assets, systems, rules or training? Briefly describe the required new infrastructures, including where possible, any requirements for successful implementation (e.g., incentives, funding, partnerships).

- A new and more expansive approach to data governance is required to address the regulatory challenges in this space. For example, the current plans use of frameworks for governance and regulation might be further enhanced with identification and development of enforcement and policing measures that ensure governance and regulation are impactful. Currently, many well-meaning frameworks exist, without the means to enforce their principles. The explosion of non-binding AI regulation and ethics documentations is a testament to this, as are the few examples of enforceable undertakings regarding AI breaches and ethical misadventures. The current plan does also not suggest governance and ethics arrangements that are responsive or agile enough to deal with emergent innovations that may present ethical concerns. A legislative solution and enforceable undertaking through the OIC, ACMA, or other government stakeholders is something worth considering and partnering with to develop.

Brokering high-value partnerships for innovation

We have no comments for this section.

Concluding remarks

We thank the Academy for this opportunity to comment, and hope our comments have been of value in supporting this important work.

Dr Ashlin Lee

Lead of the CSIRO Social Science and User Experience (SS/UX) Cross Cutting Capability, on behalf of the SS/UX Working Group

The CSIRO SS/UX Working Group is: Dr Aamir Tooba; Dr David Douglas; Dr Andreas Duesner; Dr Simon Fielke; Dr Laura Kostanski; Dr Ashlin Lee; Dr Jane Li; Mr Martijin Mooji; Dr Cara Stitzlein; Mrs Lara Tzafaris; Dr Marlien Varnfield