

Integrated Research Infrastructure for the Social Sciences (IRISS) Response to feedback on IRISS Project Plan

The IRISS project team received five submissions in response to the initial IRISS project plan. Three key themes emerged from the feedback provided - social media and real time data sources, qualitative and unstructured content, and sustainability and engagement.

1. Social Media Data

Several submissions noted the need to bring social media data into the IRISS program, as it forms a core resource for social science research. Each group notes that there is a need for the infrastructure proposed in the current plan, but that there are other core sources of data that they rely on their work, from both social media and other real-time data sources. They propose some important extensions to the IRISS platform that would enable both access to and integration with such sources, including the Australian Digital Observatory, VOSON and RAPID facilities.

As noted in our project plan, the project team has opted to exclude social media data from the first phase of the IRISS project for practical reasons, as funding and time available for the project require tight scoping of project activities. We do however acknowledge that extending the functionality of IRISS to support social media content is a key priority in any extension of project activities into the future. We also note the comment from one university that "the area of social science captures a range of approaches, needs, and constraints that benefit from a more decentralised approach". The design for IRISS has been developed to accommodate such decentralisation by focussing integration points between different systems and facilities (both current and future).

To this end, we would seek to coordinate with ARDC and the other projects to incorporate these requirements into our project activities. We would suggest three possible activities here:

1. A HASS-I program integration activity focussed on requirements gathering and technical design for social media and real time data support.
2. A design workshop at the start of the IRISS project that brought interested parties together to help define the user requirements of the infrastructure, to provide insight on what we should consider to ensure that we can scale to accommodate social media 'integration' into the future e
3. A future developments workshop in Q5-6 of the project to 'position' the community for a more expansive and inclusive infrastructure capability, leading into a proposal for HASS-I and IRISS Phase Two

One submission asked regarding whether a preferred provider has been selected for social media infrastructure development. IRISS has not identified any specific partners in this first phase of activity. We would strongly support including interested established groups working in this area to engage with in further planning in this area.

2. Qualitative and unstructured content

One submission highlighted the need to support qualitative and unstructured content, including document, image, audio and video collections. They also highlight that there are investments occurring that may facilitate some of this support - where "investments in the Linguistics Data Commons of Australia, the Australian Text Analytics Platform and the Australian Digital Observatory ensure Australia is not left behind in the capacity to undertake basic audio and textual analysis." The submission goes on to note that "...it is essential that we build on and extend this base into more sophisticated data collections and analytical capabilities "

We would acknowledge again the need for support for unstructured content and qualitative data sources in the IRISS platform. This again is limited somewhat by project timeframes and resources. We do expect however that the capabilities being established within the Australian Text Analytics Platform will be amenable to the forms of analysis suggested in the submission. To this end, the Census data demonstration project is intended to pilot some of these questions, to leverage tools within IRISS and ATAP to process structured and unstructured content in Australian census publications.

3. Sustainability and engagement

One submission highlighted the need for coordination and integration of activities both across HASS-I projects, and with other existing infrastructure. They note that "current project plans mostly build upon current systems (apart from the Trove project) affiliated with a small number of institutions for a few targeted researchers. Indication of a pathway to scalability would be beneficial."

The project brings together five institutions for the first phase of this project, that have a deliberate focus on scalable solutions. Between the partners, there is a user base of 15,000 users for AURIN and 7,500 users for ADA, highlighting that scalability considerations are already embedded in the planning of project partners.

The contributor also suggested additional communication around sustainability, cultural protocols and intellectual property. The project partners involved in IRISS have long-standing records in both research activity and research infrastructure provision. ADA, AURIN and ACSPRI are all infrastructure providers, and have demonstrated long-term sustainability (up to 40 years in the case of the Australian Data Archive). With regard to cultural protocols, the IRISS project will work closely with the Indigenous Data Network on aligning practices with the indigenous governance framework to be established in that project, and Len Smith from IDN will hold a position on the IRISS steering committee. In terms of intellectual property, ADA and AURIN have extensive track records of supporting research data and content consistent with the property rights of intellectual property holders. (For example, details of licensing for ADA can be found on their website here: <https://ada.edu.au/depositing-data/>). The four projects have also proposed an integration activity to look at legal and governance questions around integrated data products.

Finally, the submission also indicated that "each project plan would benefit from addressing cultural change management associated with moving to national infrastructure. Some information that would be useful for end-users includes:

- steps that will be taken to communicate

- participation opportunities,
- milestone consultation
- reporting outcomes of the projects as they are achieved, and
- the shift to use of national infrastructure"

Within IRISS, these engagement activities will form part of the IRISS project management work package. This engagement will occur in three ways:

1. As noted in the project deliverables, each work package has an initial consultation phase around requirements which allows for user and community input
2. Outcomes of the project will be reported through various channels including formal reporting, conference and community meetings (both in eResearch and discipline-specific fora)
3. The project steering committee will provide representation across a broad spectrum of the IRISS user community, including academia, government and the non-profit sector

There is also a proposed integration activity to support training and engagement that would assist in this engagement process.

On a related note, a related submission note that there are international services and infrastructures that could also be connected. They note that "the IRISS platform will need to be extended to draw on and be interoperable with international platforms including data commons such as the European Open Science Cloud (EOSC) and the Social Sciences and Humanities Open Cloud (SSHOC), Research graph, Wikidata and Wikipedia". The approach we have proposed is intended to support this integration, with the VASSAL and SPIRE work packages directly oriented to connecting to these international data commons, through services such as the CESSDA Vocabulary Service.

4. Other feedback

One contributor raised questions about support for transport surveys and music data. We have spoken directly with this researcher, and have proposed working to support the transport surveys in collaboration with AURIN (hosts of the Australian Transport Research Cloud). ADA already holds a number of transport surveys and can support this content. Music data falls outside of the IRISS remit, but should be considered for future investments in cultural collections.