



HASS Research Data Commons

Australian Historical Association Submission

The Australian Historical Association (AHA) is the peak national body of 850 historians and students, which includes academic, professional and other historians working in all fields of history within multiple contexts. Our organisation is committed to quality and integrity in the practice of history and advocacy on issues relating to the history profession in Australia. We count among our members many historians with deep expertise in digital research data, including leadership of major Australian Research Council projects.

The AHA welcomes the opportunity to explore ways that the current HASS Research Data Commons proposal moves further into areas that would provide significant benefits for researchers and members of the public alike.

All four projects are of interest to AHA members – for example, the Linguistics Data Commons will develop text analysis tools that could be used on historical documents, the social science project will improve access to historical census data, and the Indigenous Data Network will be developing a portal to Indigenous data and collections. However, the Trove project is the focus of our submission, especially on elements of consultation and collaboration that are currently missing.

Do you think this proposal meets the requirements?

- The evaluation criteria cite the need for ‘appropriate governance planning’. The only governance arrangements described in the draft plan refer to the existing Trove Strategic Advisory Committee which is composed of representatives from Trove’s partner organisations. The needs and opinions of Trove partners, who contribute data to Trove, are quite distinct from those of researchers making use of that data. A more suitable governance structure would include representatives from data providers, researchers, and independent technical experts.
- Plans are expected to ‘enable collaboration and build communities’ and demonstrate ‘research leadership’. However, engagement with the research community seems limited to the opportunity to beta test the close-to-final product. To ensure that this project builds research capacity, supports the development of new methodologies, and attracts a community of users, researchers need to be more actively engaged with the planning, development, testing, and release phases of the project.
- Plans are expected to develop ‘collaborative tools’, and use ‘shared underlying infrastructure’. However, no partner organisations are involved in this project, which seems solely aimed at developing a new stand-alone facility within the Trove application. A greater focus on collaboration and integration would identify overlaps with existing tools, and create opportunities for continuing, sustainable development.

- Researchers need some confidence that the tools they base their projects on will continue to be available. This plan offers no strategy for sustainability and indicates the new services might not be continued beyond the project period if further funding is not provided. A more creative and collaborative approach is required.
- As such, the development of a platform that would enable more high-level computational analyses of its data has inherent potential for historical researchers, especially as most do not possess advanced technical capabilities or coding expertise. However, our members are concerned that the proposals as currently outlined are of limited utility in advancing existing analytical tools.
- The proposals seem to replicate tools already available through the popular platform [GLAM workbench](#), which can likewise be used to create visualisations based on Trove search queries without any need for coding knowledge or advanced technical abilities on the part of researchers. If the proposal is essentially to integrate the architecture of the workbench into the Trove website itself, this should be more clearly articulated.
- The project proposal focuses on the question of data visualisation, rather than data analysis. This raises concerns about whether these tools will be capable of driving new research queries and thereby lead to widespread take-up among researchers. A tool for mapping results of an individual search query on to a map or plotting it on a time graph would definitely be of use to historians wanting to engage the public with their research questions or as an easy way for visually displaying the results of their queries. Its ability to deliver new data and thus contribute the fresh insights that in turn generate journal articles, books and research project proposals is less clear.
- To ensure that this proposed development does deliver new capabilities for driving original research, we suggest that the platform should ideally: 1) support more robust forms of quantitative analysis, such as by returning measures of the geographic spread of material retrieved in a search query; and 2) support analysis of large-scale data by enabling comparison of results from more than one query.

How would this proposal support your research?

- Trove has changed the practice of history in Australia, but most users of Trove employ traditional research methods. This proposal offers the opportunity to support the development of new research methods that examine historical resources at scale. However, for this to be really useful, it needs to be part of a continuing, and responsive, program of collaboration with researchers, and not a one-off product.
- The AHA has a number of members vitally interested in extending the Trove information sharing framework and therefore want to see this proposal extended further and developed to support research. For example, Professor Mark Finnane (Griffith) and Dr Alana Piper (UTS) have examples of local shared digital platforms ([Criminal Characters](#), Prosecution Project) developed with researcher and crowd-sourced support. What they suggest is something more generic, such as a user-built national depository of information in archive and manuscript collections that would standardise data retrieval through use of proper metadata standards and so increase the possibilities of indexing and retrieval of the micro-data within documents.
- Another potential development for advancing research by leveraging Trove data would be to integrate qualitative data tools into the site, such as for topic modelling and sentiment analysis. A model for how this can be done to enable even those with limited technical skills to conduct such analyses is provided by Gale Cengage, whose [Gale Digital Scholar Lab](#) now enables data mining of their various corpora. The ability to apply

similar tools to Trove data would be of tremendous value not only to historians, but to scholars across the humanities and social sciences.

Are there other capability gaps that we should consider in the longer term?

- There is no doubt that Trove is a core piece of national research infrastructure and we hope that this will be reflected in future Research Infrastructure Roadmaps and Investment Plans.
- The resources to be analysed by this new facility are limited to digitised books, periodicals, newspapers, and gazettes. Tools and services to support research using archives, maps, images, and web archives are also needed.
- The creation of new tools and platforms needs to be accompanied by plans and investments for the development of digital research skills.
- Trove cannot be the sole focus of these investments as large quantities of valuable research materials are held by other institutions. Tools should be developed that work across, rather than within, existing collection boundaries.
- Investment in digitisation, available to a range of cultural heritage and research organisations, needs to be recognised as a core component of HASS research infrastructure.
- Development of a platform to advance data discovery and sharing in historical archives and manuscript collections. Digital information infrastructure has the potential to transform this research field. Innovative web-based systems like [Zotero](#) provide a model for advancing information discovery, storage and access on a shared basis, multiplying potential benefits. Crowd-sourcing of information for building web-based data repositories highlights the potential of the data sharing approach (see eg [Atlas of Living Australia](#)). Development of an Australian platform to advance data discovery and sharing in historical archives and manuscript collections would encourage current and future researchers in Australian humanities studies (history, literature, politics and so on) to access a dramatically expanded national resource.
- While data visualisation tools could play an important role in this work, what would be more significant to most are the possibilities for data analysis that this platform could open up. We would therefore urge strong consideration of different research use cases aimed at analysing, rather than just visualising, Trove data.

27 September 2021