



# Data and Services Summit report 2019

29 November 2019

## Summary

The ARDC Data and Services Summit provided an opportunity for collaboration between institutions, research communities and infrastructure providers, who together identified some of Australia's major research data opportunities and challenges and worked towards shared solutions.

The [Summit](#), which took place on 21 October 2019, was co-located with the ARDC Infrastructure Summit and eResearch Australasia conference in Brisbane. Attended by delegates from across the sector, the Summit reflected on the learnings and outcomes of the 42 Data and Services Discovery projects funded by the ARDC. These projects covered a range of disciplines, including health and medical sciences, imaging and characterisation, life and natural sciences, and the humanities, arts and social sciences. The summit also considered how future directions and initiatives might be effectively aligned from the perspective of institutions, research communities and infrastructure providers.



The ARDC outlined its proposed approach to providing data infrastructure over the next three years, emphasising the importance of strategic partnerships with the sector in a balanced portfolio of projects. Adrian Burton, Director Data, Policy and Services outlined a set of six proposed programs in which the ARDC would collaborate

with research institutions, infrastructure providers, industry and the public sector to develop national data assets that support leading edge research. Delegates also heard about future directions from other national data infrastructure providers, Bioplatforms Australia and the Australian Data Archive, as well as the Research Data Culture group (a collaborative working group representing enterprise services at several universities).

## Sector-wide challenges and shared learnings

The following themes emerged across the day's discussions:

- Sensitive data including: health, personal, cultural, environmental, and commercial. Delegates identified:
  - Access control as a common service needed across all domains working with sensitive data.
  - The importance of trust when asking providers to share data; where trust is uncertain, providers default to closing access to their data. Many approaches to building trust were discussed: building relationships within a community; guarding closely against data loss and misuse; and developing tools to help providers categorise, identify, track and control that risk.
- Sustainability of services. Delegates described and discussed:
  - The difficulty of transitioning one-off projects into ongoing, “business as usual” services.
  - The importance of adopting existing solutions to share maintenance burden; formalising relationships with institutions and stakeholders and clearly defining roles and value proposition in partnership with those stakeholders.
  - Demonstrating impact to ensure attractiveness of continued funding.
- Standards are vital to support data sharing and reuse. In some domains (e.g. imaging and characterisation), delegates discussed how metadata creation could be automated at the stage of data collection to make interoperability easier to achieve; in others (e.g. health, agriculture, water and energy), delegates observed that data collected from sources where research is not the primary purpose is unavoidably heterogeneous in format, and requires additional support for standardisation. Across all domains, delegates discussed the importance of communities of trust coming together to define standards.
- Data growth is a concern across the sector - data is growing faster than storage efficiency, leading to ever-rising management costs. Delegates discussed the importance of archival practice (appraisal, retention, disposal) and curation (validating, cleaning, processing) from the perspective of both storage providers and researchers.
- Multi-party collaboration was identified as both a necessary but challenging component of valuable data assets for research. The summit saw such collaboration as vital for scaling data resources, pooling data capability, and enabling sustainability and standards. However, delegates noted that natural inertia and competitive forces often pull in the other direction.

## The role of the ARDC

Delegates gave feedback about the role of the ARDC in addressing these issues. Suggestions included providing guidance about best practice in data management and sharing, supporting key services such as persistent identifiers, and helping to coordinate efforts across the sector. In particular, delegates expressed the importance of the ARDC's role as a trusted partner that develops strategic partnerships and facilitates cooperation between stakeholders. This could be achieved both at a local level, by bringing together communities of practice and hosting intensive problem solving events, and also at an international level by leveraging relationships with other overarching data initiatives and bodies. The assembly suggested that some vital elements for effective research

data (collaboration, community action, standards, sustainability, strategic focus) would not emerge naturally or quickly enough from the research system without some catalyst from a cross-cutting sector-wide national body like the ARDC.

## Benefits and outcomes

The summit provided a forum for the national conversation on the role of data in leading edge research. This public discourse made it possible to identify shared issues and opportunities for joined-up effort. It also represented an opportunity to consider a broader framework of complementary and coherent data activities among institutions, research communities, and national infrastructure providers.



### Breakout Discussion: “What are your most important challenges/pain points?”

The design of the Summit allowed for national conversations both in specific domain areas (such as Humanities or Medical research) as well as across and among domains, institutions, and national infrastructures. The Summit:

- Increased communication and collaboration in domain areas and among institutions.
- Identified important ongoing challenges (sensitive data, sustainability, standards, data growth, complex collaborations). These have been incorporated into the refreshed strategy and focus for ARDC.
- Identified specific opportunities for ARDC to play a complementary role to research communities and institutions (strategic partnerships, communities of practice, guidance, coordination). These have been incorporated into ARDC project, services, infrastructure, and outreach planning.
- Identified areas where continued co-design of a national data system would be valuable:
  - Data culture, infrastructure, policy and support at enterprise level in research institutions and their role in a data commons;
  - Domains and disciplines moving to national data approaches
  - Business and public sector touchpoints

The ARDC will progress these national agendas over the next year.