The ARDC

Evolution of a National Research Data Infrastructure

PRESENTED BY

Ian Duncan
(Director, Outreach)

ACKNOWLEDGEMENT OF COUNTRY

We acknowledge and celebrate the traditional owners and custodians of the lands on which we meet and work, and we pay our respects to elders past, present, and emerging.
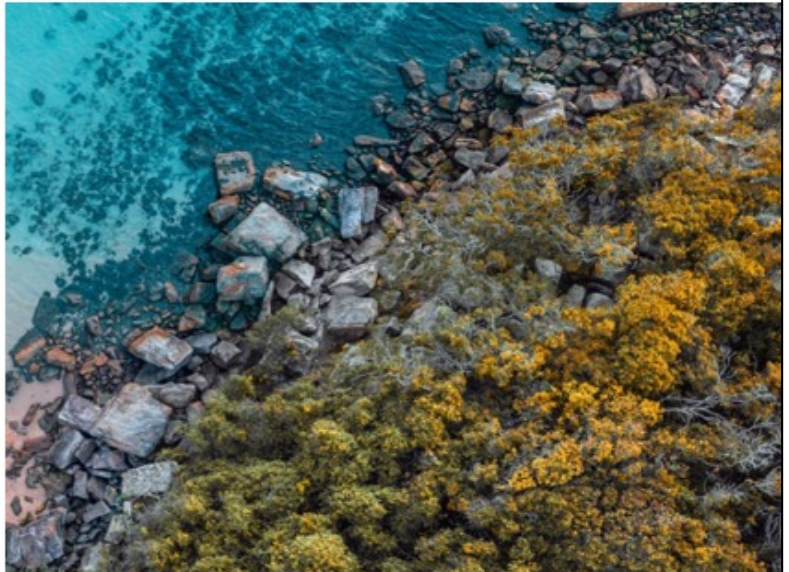
Image — Will Turner - qtWiw4MdF5Q / Unsplash.com

ARDC acknowledges the Traditional Owners of the lands on which we live and work across Australia. We pay our respect to Elders past, present and emerging and we recognise their valuable contributions to Australian and global society.

Australia is an established global leader in world-class research. The Australian Government helps maintain this reputation by ensuring researchers have access to cutting edge national research infrastructure through the NCRIS program.

The ARDC was established under the Australian Government's National Collaborative Research Infrastructure Strategy (NCRIS) in 2019 and was created from three earlier data-focused national infrastructure projects: Australian National Data Service, Nectar, and Research Data Services.

- NCRIS currently supports 22 funded projects, plus pilot projects and international membership of the Research Data Alliance. The projects are led by universities, publicly funded research organisations as well as private and not for profit companies.
- This national research infrastructure could be a physical instrument like an electronic microscope, or a service like manufacturing of nano technology, or digital like the cloud compute and tends to be bigger in scale than any single or group of universities could provide their researchers.
- The projects form a network involving over 200 delivery partnerships, and employ over 1900 highly skilled technical experts, researchers and facility managers.
- Projects provide merit-based access for all Australian researchers to their infrastructure with over 65,000 Australian and 12,000 international users supported every year and a 94% user satisfaction rating.
- They also have high co-investment rates. As of the 2018 census, each $1 Government invested in NCRIS saw $1.29 in co-investment from universities, research agencies, state and territory governments and industry.

A key aim of the integration of ANDS, Nectar and RDS was to ensure that the digital research ecosystem remained accessible and responsive to the broadest possible range of researchers and research priorities. This was delivered through a new corporate structure based on a foundation of research institution membership as well as engagement at several levels within those organisations to identify priorities and challenges.

The ARDC now has over 70 staff based at host institutions around Australia and 20 member institutions and I will be outlining the success we have had in increasing the breadth of participation across all of our programs shortly.
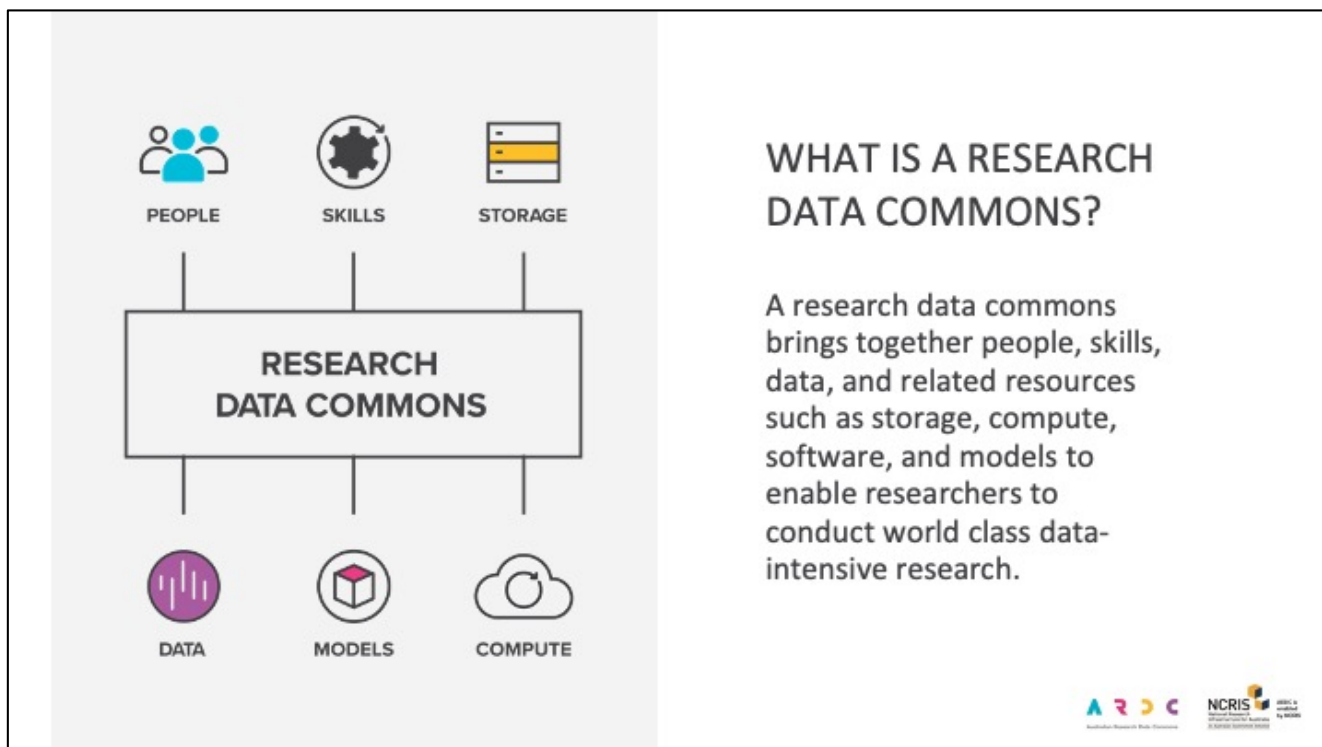
The original three organisations addressed different and complementary slices of the digital research challenge facing researchers and another key aim of the integration was to link those parts together to create simpler access, stronger links, and a coherent suite of offerings to individuals and organisations in creating a competitive advantage for researchers through data.

So what is a research data commons?

A research data commons brings together people, skills, data, and related resources such as storage, compute, software, and models to enable researchers to conduct world class data-intensive research.

We use the data commons model to describe how our activities link together and while we are called the Australian Research Data Commons, we're working together with our partners to build a national research data commons for Australia.

We support open science, and specifically the FAIR principles which ensure that data is findable, accessible, interoperable and reusable for all researchers.
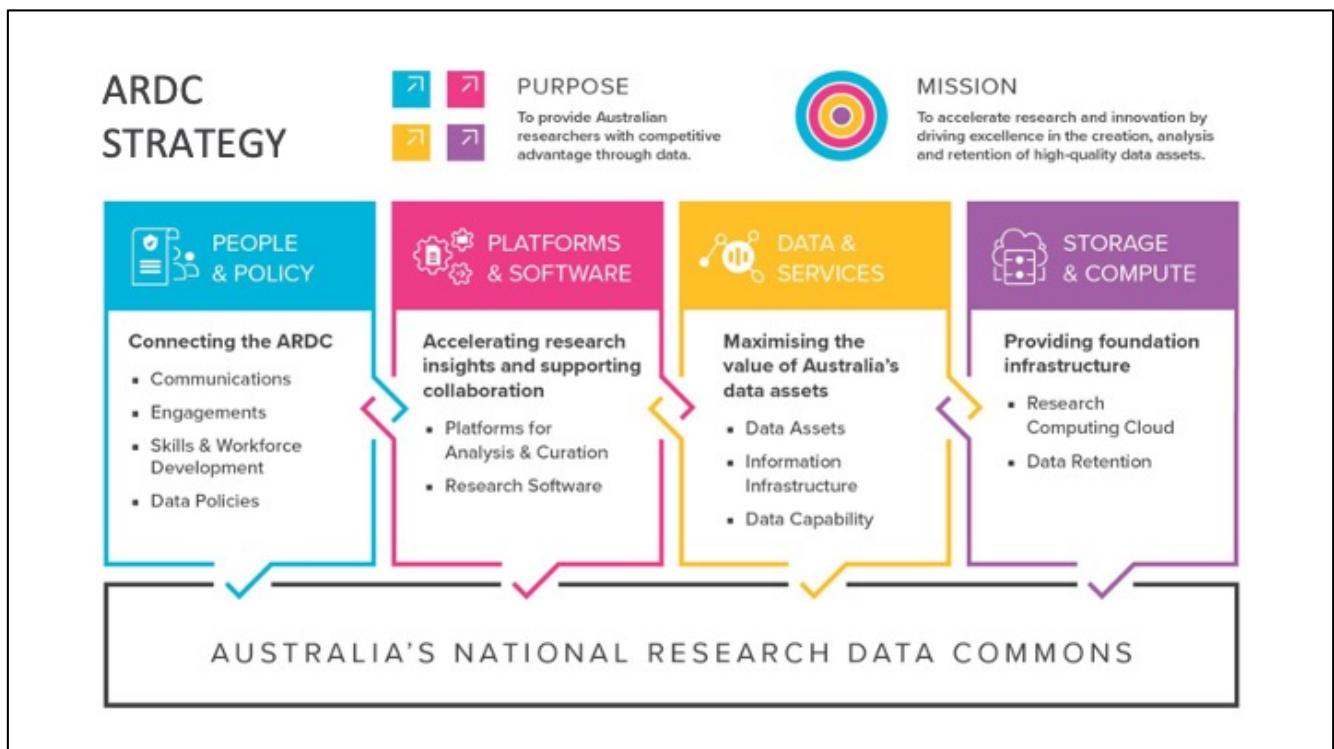
As I mentioned on the last slide, the ARDC's purpose is to provide Australian researchers with competitive advantage through data and we aim to accelerate research and innovation by driving excellence in the creation, analysis and retention of high-quality data assets.

Over the last two years we have run a range of open calls that have given us a unique insight into the digital research infrastructure needs of the Australian research community.

This slide shows the current ARDC strategy on a page and highlights our four portfolios, through which we operate our range of programs and services.

- Communities including ML4AU, machine learning, Research software, Geospatial data
- platform projects that draw together research data, models, analysis tools and workflows to support collaborative research across institutional and discipline boundaries
- national data collections that create or develop high quality data collections that support leading edge research and are national in scale
- Services including nectar, rda, rva, and persistent identifiers

The ARDC skills and workforce development programs create national tools and materials for researchers and trainers that make discovering and accessing the right resources simpler and more flexible.

The Digital Research Skills Australasia portal is a combined project funded by the ARDC and the Pawsey Supercomputing center and including 12 partners to make it easy to find DRI skills materials, events, and trainers across Australian and New Zealand (Dresa is available to NZ providers and researchers via Tuakiri)

The ARDC also developed a consortium model for access to the Carpentries model and the ARDC Platinum Carpentries membership gives research institutions of all sizes in Australia access to instructor training for The Carpentries at a standardised reduced rate.
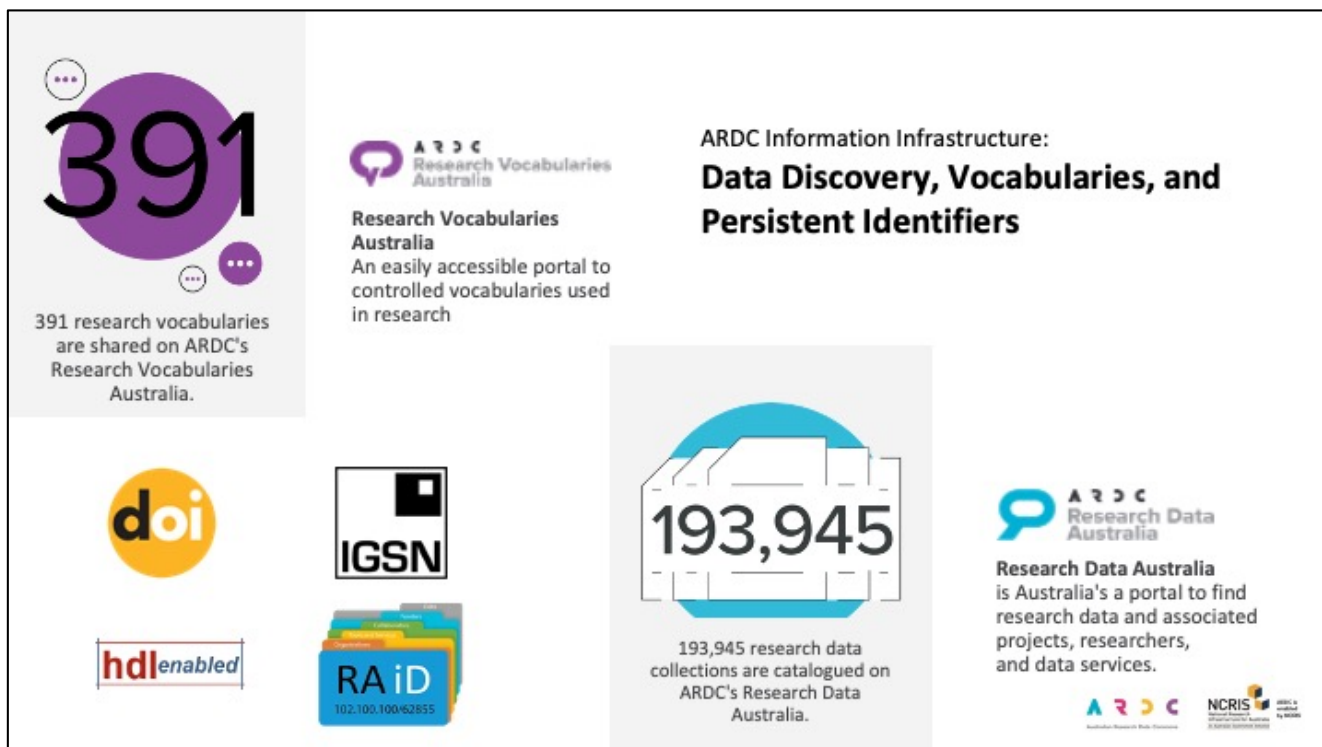
The national information infrastructures of Research Data Australia, Research Vocabularies Australia, and our Persistent Identifier programs and services have been highly successful and continue to evolve as a suite of coordinated services that provide the fundamental identification, discovery, and cross-discipline services needed from a coordinated national data commons.

Our region has been a leader in adopting and developing approaches to persistent identifiers and we intend to continue that emphasis on these core services and tools.
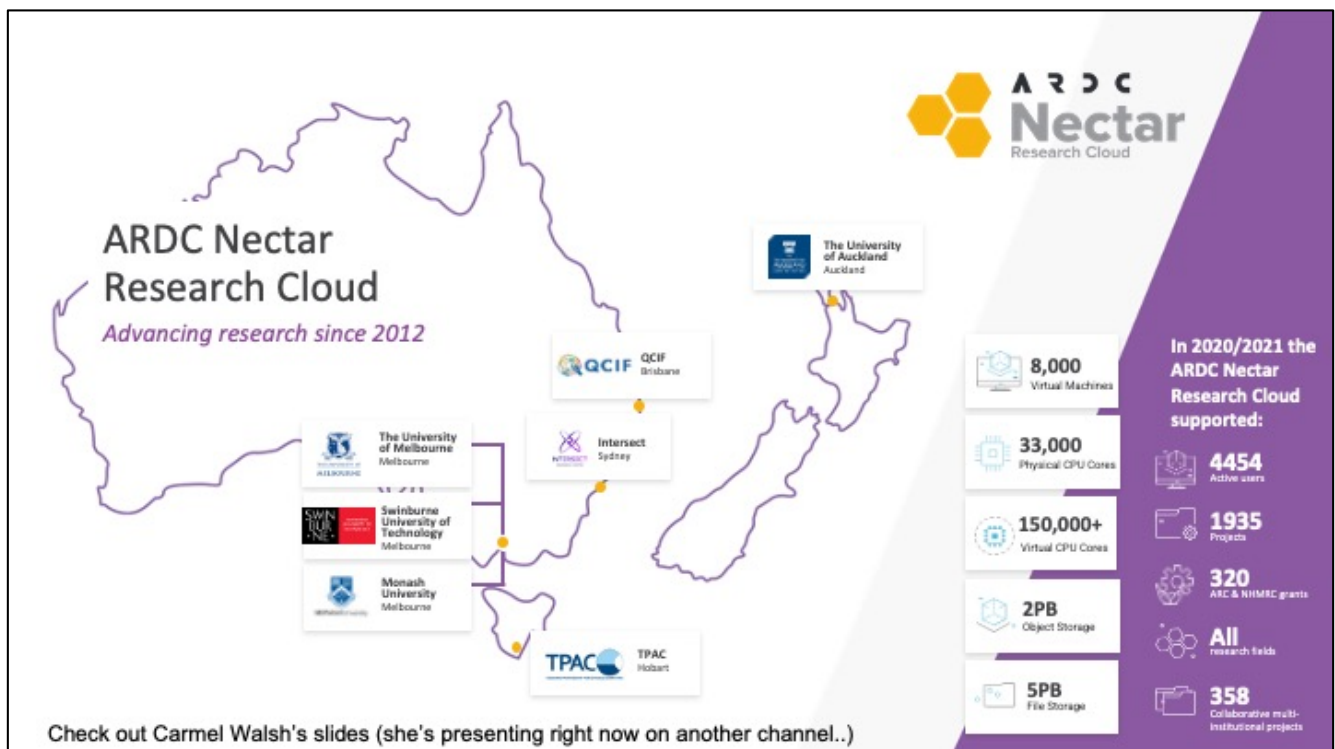
Our platforms and data assets need somewhere to live and somewhere to operate and a significant development in our strategy has been ensuring that our underpinning infrastructures are also evolving to meet these new challenges and operating environments.

The Nectar research cloud has been extremely successful and is a highly prized resource for researchers. Additional investment in 2019 has ensured researchers have agile compute resources available and the strategy for the Nectar cloud is continuously evolving to meet the needs of both researchers and our partners who operate the cloud (including of course the University of Auckland)

Allied with the Nectar research cloud is the ARDC Data Retention program, which has moved away from supporting raw storage capacity, which brings its own sustainability challenges, towards testing new methods of supporting the curation and stewardship of key data assets through their FAIR characteristics.

Let's look at some examples of the benefits of our national approach:

- 25 Universities are working together to collectively create a nationally agreed institutional research data management framework under the Institutional Underpinnings program
- There are 15 Communities of Practice including areas like machine learning, research software, and sensitive data
- There are 26 national research infrastructure and data platforms
- Research Data Australia supports nearly 50,000 unique searches per year
- The Nectar Research Cloud supported 1788 projects last year

These statistics highlight how the national data and research infrastructure capabilities of ARDC are extensively utilised across the sector and how they foster national collaborations to deliver research impact

But there is still significant unmet demand for platforms, data assets, underpinning infrastructures, and services, as well as the challenge of increased translation of research into industry.

We have therefore decided to transform the delivery of our national research infrastructure through a more strategic and coordinated approach, to ensure maximum benefit to the maximum number of researchers across all sectors.

In developing this new approach we are testing several potential models for our future activities.

HEALTH STUDIES AUSTRALIAN NATIONAL DATA ASSET
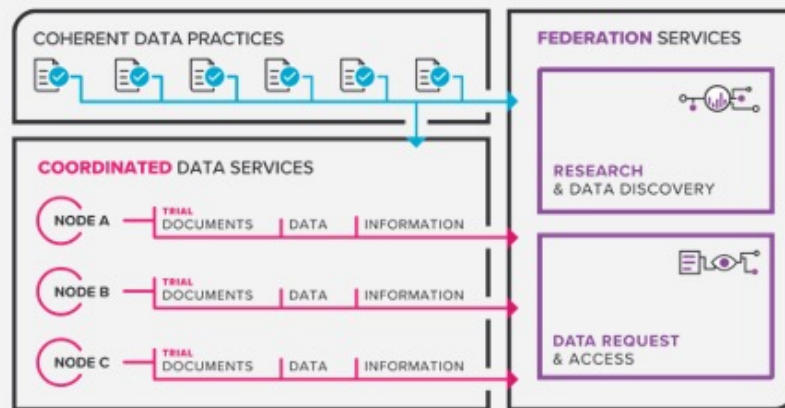
The first focuses on data and is the Health Studies Australian National Data Asset, shortened to the acronym HeSANDA.

The HeSANDA program aims to build a national infrastructure that can support the sharing and reuse of health research data and increase research impact and integrity.

While the potential scope of HeSANDA is huge, in the initial phase we are creating a national data asset that concentrates on building a collaborative and shared approach to NHMRC supported Clinical Trials and Cohort Studies. This program now includes 91 collaborators from 72 organisations, including 18 universities, 10 medical research institutes, 19 health service operators and 16 clinical trials networks.
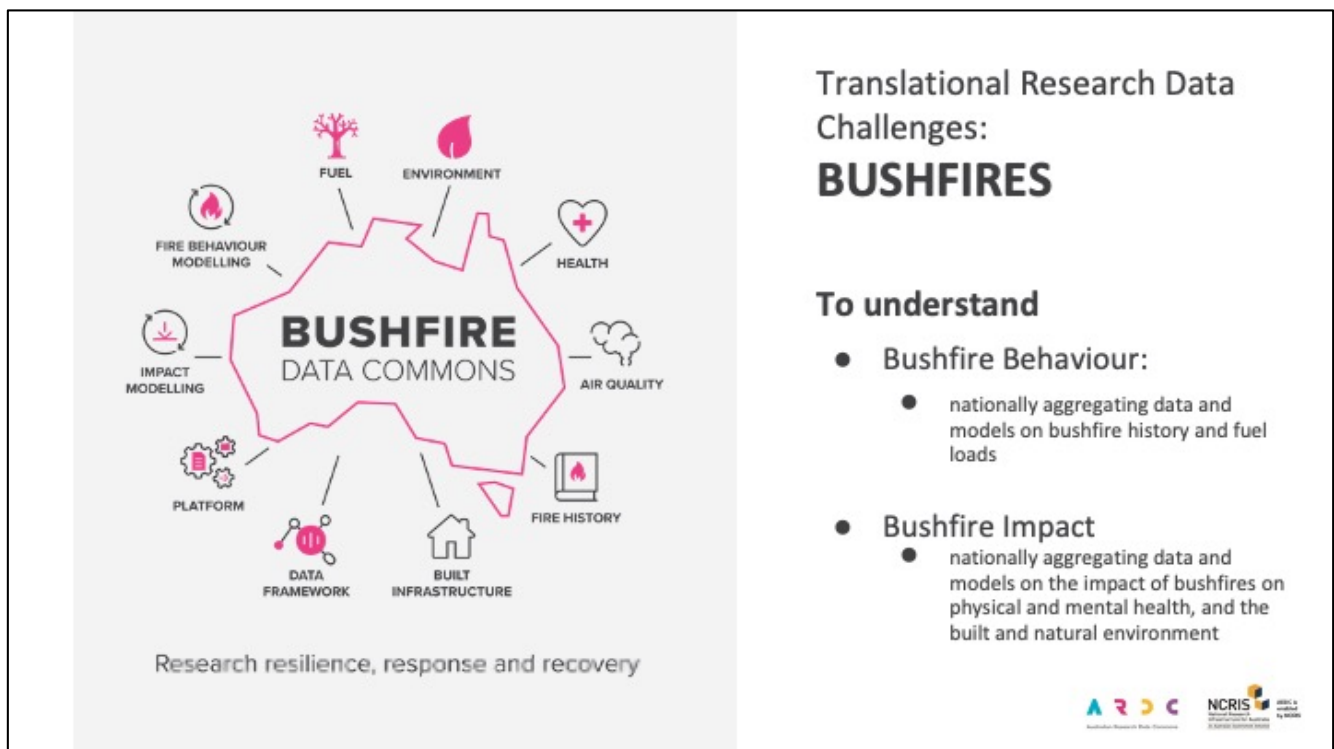
The second model sits in our Translational Research Data Challenges program and concentrates on a national approach to Bushfire Data Infrastructure including data and tools and digs into the challenges of creating an output which spans government, industry and the research sector.

To protect Australian citizens from the dangers of bushfires, firefighters and planners need to be able to model and predict bushfire behaviour. For example, allocating bushfire responders or planning prescribed burning and urban development. The Bushfire Data Commons aims to build comprehensive national data assets in these areas and create a shared bushfire modelling environment that will support research, development, planning and response.

In designing this program ARDC conducted 9 months of targeted consultations, including a series of facilitation meetings and design workshops with 92 stakeholders, which identified the key capabilities required.
● The first is understanding the behaviour of bushfires themselves, aggregating models and data of bushfire history and fuel loads.
● The second key area is developing an understanding of the impact of bushfires on factors including population health (both physical & mental), smoke and the climate, flora and fauna, and urban infrastructure.
● This program also dovetails with the national bushfire intelligence capability.
● There are many agencies involved in bushfire research and operational response, working with the ARDC in the definition and delivery of the data challenges and reinforcing the benefits of ARDC's unique positioning in being able to bring major cross-sector collaborations together.

Humanities, Arts & Social Sciences Research Data Commons and Indigenous Data Capability (HASSi-RDC)

The third initiative is the Humanities, Arts and Social Sciences Research Data Commons (or HASS RDC) and Indigenous Data Capability which investigates how distinct and discipline-specific projects can be guided in such as way as to fit together to create a suite of products which meet the needs of discrete communities while opening the door to new collaborations, increased efficiency, and improved sustainability.

The four activities in the HASS RDC were made possible through an additional funding round in October 2020 following an analysis of potential areas of activity by ARDC and by acting as a coordinating driver for the projects, the ARDC can engage deeply with the relevant communities to identify and build capability and linkages and ensure that best practice is shared across the projects and to other domains and disciplines.

A key consideration of this process of coordination is identifying activities that will benefit all four streams. One example is the benefit from standard access and authentication to the resulting outputs for research communities, which will facilitate future collaborations. We anticipate that all four streams will leverage this shared access and authentication framework which can then be rolled out beyond the HASS RDC to our other programs.
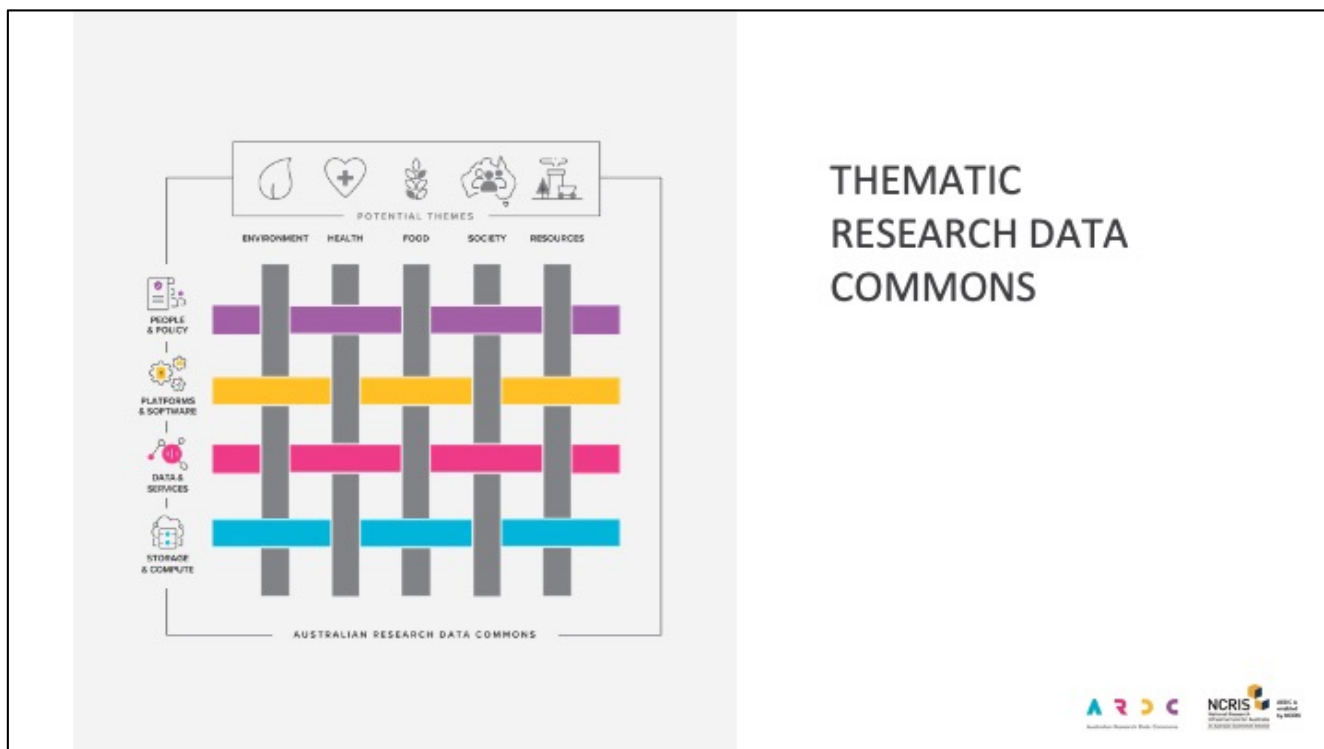
Following lessons learnt from the models described earlier, the ARDC's future vision is based around the concept of Thematic Research Data Commons - environments that enable us to support the maximum number of researchers through a smaller number of strategically selected priority areas.

Essentially, a fabric of NRI capabilities selected strategically rather than competitively and codesigned and delivered with the research community

The fabric has both nationally focussed platform capabilities that strengthen and support the broader system – the horizontals - and a deep focus on identified national challenges and opportunities – the verticals - to provide an ideal balance for the national system.

As a hub of expertise, the ARDC is positioned to drive best practice in the creation, analysis and retention of high-quality data assets and share this expertise across domains
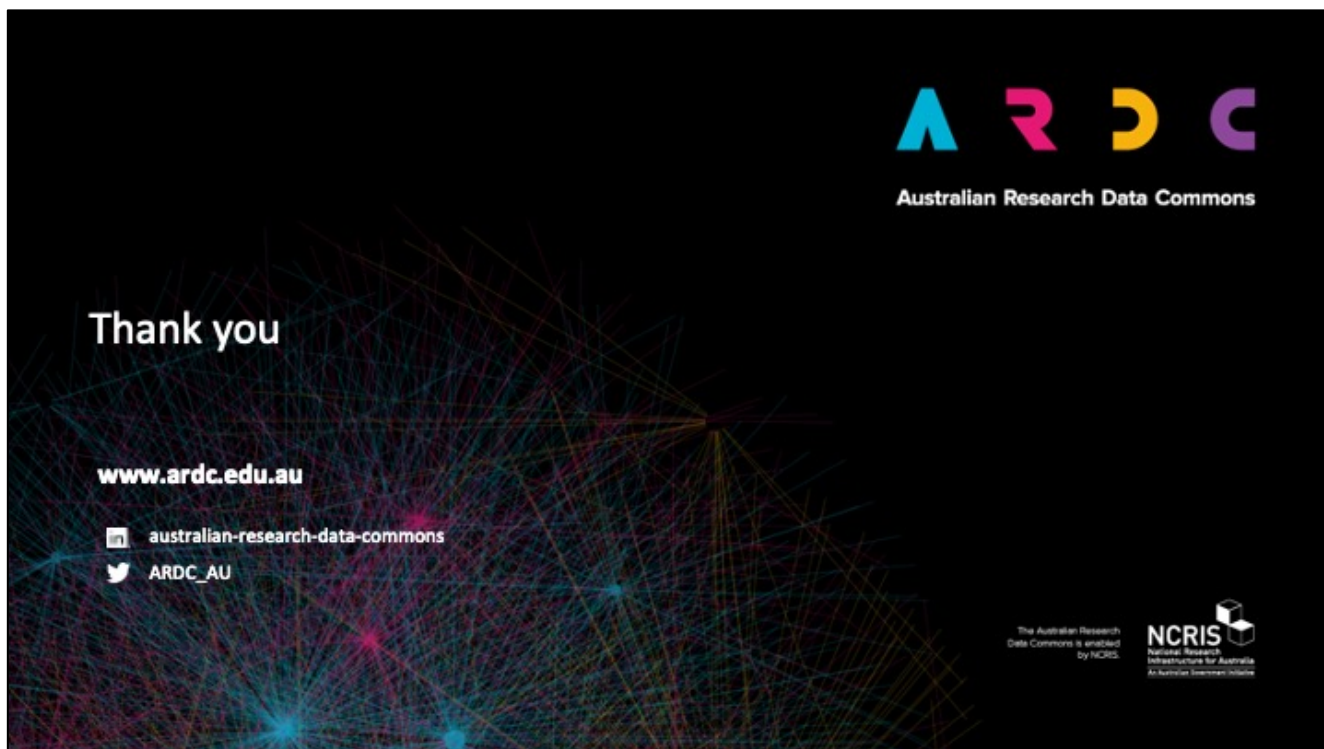
In closing, we look forward to partnering with you in developing the strategy and implementing the Thematic Research Data Commons.

If you'd like to stay in touch with ARDC please subscribe to our newsletter on the website and follow us on twitter and LinkedIn

Thank you and I'm happy to take any questions.