

# ARDC Submission on the Regional Data Hub

The Australian Research Data Commons (ARDC) welcomes the Regional Data Hub Consultation and federal funding to support the construction of this Hub under the Better Data Use to Support Delivery for Regional Australians program. As noted in the consultation paper,

data plays a critical role ... [d]ata enables better, faster and more efficient design and delivery of programs, policy and services to meet the needs of communities. Easy access to the right data can help people and communities have their say on local needs, plan investments and make decisions.<sup>1</sup>

In this paper, we offer comments on issues raised in the consultation paper and respond to selected questions drawing upon our expertise and experience working in the field of digital research infrastructure and research data sharing. Specifically, we will comment on the research sector and suggest contact organisations. In addition, we will explain the differences between open data and FAIR data to comment on the proposed design of the Regional Data Hub (or "Hub").

## I. Researchers and Research Organisations as Key Stakeholders

We note that the consultation paper identifies "researchers and consultants" as a category of stakeholders with "an interest in regional issues."<sup>2</sup> While we agree with this observation, we would add that universities and research organisations are also stakeholders which the Bureau of Communications, Arts and Regional Research (BCARR) would "want to hear from."<sup>3</sup> In general, researchers do not operate on their own but rather as employees of teaching and/or research institutions. They rely upon their employers for institutional support services including access to IT infrastructure such as online data repositories. For example, the CSIRO operates a Data Access Portal which enables their research staff to retrieve data, software and other digital assets collected through their diverse research activities. This resource would likely hold some data sets of interest to regional stakeholders and communities. Similarly, there will likely be data collections on regional Australia stored in digital repositories established at Australian universities. It follows that researchers, universities and other research organisations are "data owners" or "data holders" who may be in a position to contribute data to the Regional Data Hub.<sup>4</sup> Further, given their training and expertise, researchers may be able to assist regional stakeholders and communities to leverage and maximise the benefits that can be derived from data sets acquired through the Hub. For these reasons, ARDC recommends that BCARR consider engaging with stakeholders from the research sectors. It is likely that researchers and research organisations with relevant expertise will be in a position to contribute to the co-design of the Regional Data Hub.

<sup>2</sup> Ibid 4.

<sup>&</sup>lt;sup>1</sup> Department of Infrastructure, Transport, Regional Development and Communications, *Consultation Paper - Regional Data Hub: Understanding and responding to regional data needs* (May 2021) <a href="https://www.communications.gov.au/have-your-say/regional-data-hub-consultation">https://www.communications.gov.au/have-your-say/regional-data-hub-consultation</a>.

<sup>&</sup>lt;sup>3</sup> Ibid 4.

<sup>&</sup>lt;sup>4</sup> Ibid 4.

As part of our program of work, the ARDC has commissioned a study to analyse how research data is utilised by researchers to generate social, economic, environmental and cultural benefits for non-academic end-users of research. Our study was conducted by scholars from the Institute for Methods Innovation at Warwick University. This work resulted in the publication of a two-part report, Investigating the Link between Research Data and Impact (Phases I and II, 2019 and 2020, respectively).<sup>5</sup> Our study showed that research data alone is not sufficient to create positive impacts. Strong interventions such as analysis, curation and product development are required to leverage the broader value of data for business, government, professionals, community groups and members of the general public. The study also highlighted the importance of active planning, coordination and collaboration between researchers and stakeholders to achieve impact through research data. Specifically, closer links between universities, government and industry are needed to realise the benefit of research data as a managed and shared resource across sectors. The results from this study have informed our work on the development of resources of benefit for researchers and the wider community. It is likely that our commissioned study will offer insights for work on the development of the Regional Data Hub.

# II. Data Networks and Data Infrastructure Providers as Partners

In Table 2, the consultation paper listed the task of "locating relevant data" as one of the challenges in setting up the Regional Data Hub.<sup>6</sup> Aside from universities and research organisations, we suggest that BCARR consider reaching out to the ARDC and other National Collaborative Research Infrastructure Strategy (NCRIS) facilities.<sup>7</sup> The ARDC operates Research Data Australia (RDA), a data discovery service for researchers. We will describe the work of RDA in our answers to one set of questions from the consultation paper below.

Regarding data networks, we note that the University of Melbourne established the Indigenous Data Network (IDN) in 2017. ARDC and IDN collaborated on a Data and Services Project at that time. The objective of the IDN is summarised by its motto, "putting indigenous data back in community hands."<sup>8</sup> The IDN works closely with indigenous communities and organisations to build technical

<sup>&</sup>lt;sup>5</sup> Eric A Jensen and Mark Reed, *Investigating the Link Between Research Data and Impact* (ARDC Commissioned Report, November 2019) <https://zenodo.org/record/3543505#.YNb40CN\_Ua4>; Eric A Jensen and Mark Reed, *Investigating the Link Between Research Data and Impact* (Phase II) (ARDC Commissioned Report, September 2020) <https://zenodo.org/record/4013924#.YNb5BCN\_Ua4>. <sup>6</sup> Above n 1, 7.

<sup>&</sup>lt;sup>7</sup> The National Collaborative Research Infrastructure Strategy (NCRIS) is a funding program developed by the Department of Education, Skills and Employment (DESE) to support Australian research. It is designed to initiate, maintain and sustain National Research Infrastructure (NRI). DESE defines NRI as consisting of "facilities, equipment and resources that are needed to perform research [e.g. supercomputers and datasets, as well as the] experts needed to operate it." The majority of NRI in Australia is funded through the NCRIS program. Currently, NCRIS supports an international membership, pilot projects and 22 funded projects including the ARDC. The NCRIS program is informed by a set of key principles ("NCRIS Principles") which emphasizes collaboration, co-investment and the development of national infrastructure to "serve the research and innovation systems broadly, not just the host/funded institutions." For more information, please visit the Department's NCRIS webpage: <https://www.dese.gov.au/ncris>.

<sup>&</sup>lt;sup>8</sup> University of Melbourne, Melbourne School of Population and Global Health, *Putting Indigenous data back into community hands* (4 December 2019) <a href="https://mspgh.unimelb.edu.au/centres-institutes/centre-for-health-equity/research-group/indigenous-data-network/news-and-events/putting-indigenous-data-back-in-community-hands">https://mspgh.unimelb.edu.au/centres-institutes/centre-for-health-equity/research-group/indigenous-data-network/news-and-events/putting-indigenous-data-back-in-community-hands>.

capability and provide support for effective data management including ensuring data access, accuracy and integrity. In July 2020, the IDN received \$1.3 million from the federal Government to support work on one of the priority reforms under the new National Agreement on Closing the Gap.<sup>9</sup> Numbered four, this priority reform is titled, "Shared access to data and information at a regional level."<sup>10</sup> This initiative will enable governments to establish data projects in up to six locations across Australia. Through these projects, Aboriginal and Torres Strait Islander communities will be able to collect, access and use data - including "regional specific data" - "to help drive their own development and discussions with governments."<sup>11</sup> In addition, as part of a collaborative initiative under the 2020 Research Infrastructure Investment Plan, ARDC and IDN will collaborate to build a research data commons for indigenous studies with input from the indigenous research community.<sup>12</sup>

#### What regional data do you hold?

Currently, the RDA describes more than 178,000 data sets from over one hundred Australian research organisations, government agencies, and cultural institutions. We do not store the data itself but provide descriptions of, and links to, the data from our data publishing partners. The RDA covers a broad spectrum of research fields - across sciences, social sciences, arts and humanities. Some data sets located through the RDA search function is immediately accessible online via partner sites and free to use (subject to any licence conditions). Since the RDA collection is national in scope, it holds records on data sets pertaining to regional Australia. For example, a search for the town "Wagga Wagga" returns more than 300 results for data from a number of fields of research including Earth Sciences, Agriculture and Veterinary sciences, Environmental Sciences, Built Environment and Design, Biological Sciences, Studies in Human Society, Studies in Creative Arts, History and Archaeology, and Philosophy and Religious Studies. Other NCRIS facilities also hold relevant regional data from various scientific domains such as environmental, demographics, health science, urban planning, etc. The NCRIS facility, Australian Urban Research Infrastructure Network (AURIN), for example, operates an "online workbench with access to thousands of multidisciplinary datasets, from hundreds of data sources and analytical tools covering spatial and statistical modelling, planning and visualisation."<sup>13</sup> ARDC, AURIN and the other NCRIS facilities could be good infrastructure partners that could contribute expertise for the implementation of the Regional Data Hub initiative.

<sup>&</sup>lt;sup>9</sup> University of Melbourne, Melbourne School of Population and Global Health, *\$1.3 million Indigenous data project to transform community outcomes* (14 September 2020)

<sup>&</sup>lt;https://mspgh.unimelb.edu.au/centres-institutes/centre-for-health-equity/research-

group/indigenous-data-network/news-and-events/\$1.3-million-indigenous-data-project-to-transform-community-outcomes>.

<sup>&</sup>lt;sup>10</sup> Closing the Gap in Partnership, *Priority Reform Four - Shared access to data and information at a regional level* <a href="https://www.closingthegap.gov.au/national-agreement/national-agreement-closing-the-gap/6-priority-reform-areas/four">https://www.closingthegap.gov.au/national-agreement/national-agreement-closing-the-gap/6-priority-reform-areas/four</a>.

<sup>&</sup>lt;sup>11</sup> Ibid.

<sup>&</sup>lt;sup>12</sup> ARDC, HASS and Indigenous Research Data Commons: Investing in HASS and Indigenous research communities (2021) <a href="https://ardc.edu.au/collaborations/strategic-activities/hass-and-indigenous-research-data-commons/">https://ardc.edu.au/collaborations/strategic-activities/hass-and-indigenous-research-data-commons/</a>.

<sup>&</sup>lt;sup>13</sup> The Australian Urban Research Infrastructure Network (AURIN), *The Aurin Journey* <a href="https://aurin.org.au/about-aurin/the-aurin-journey/">https://aurin.org.au/about-aurin/the-aurin-journey/</a>.

#### What data (either your own or other data) would you like to see shared through the Hub?

The RDA is a free, user-friendly website which can be accessed by members of the general public in Australia and abroad. While it is designed specifically for researchers, it has broader relevance for other groups of data users including educators, students, policy makers and business people. As the above search example on Wagga Wagga illustrates, the RDA site holds records on diverse data sets including information on many aspects of regional centres in Australia. In the RDA website, each dataset is linked as a visual graph to funding grants, organisations, researchers and published research in which the data is referenced. Data providers can therefore immediately know how their data is being used, the organisations using it and the research it contributed to producing. Additionally, the RDA site enables users to identify any published research that is linked to data that originated from their region.

Regional data users could access the value of RDA either through the existing ARDC portal, through the proposed Regional Data Hub, or via a range of other metadata aggregators such as Google. ARDC would be keen to work with contacts at the Regional Data Hub to enable researchers and other data users to find and access data required to address regional issues. This would be done by registering descriptions of data holdings from the Hub in our infrastructure and promoting research usage among regional communities around the country.

#### What would be the benefits of greater data sharing for you or your organisation?

The role of the ARDC is to provide Australian researchers with competitive advantage through data. For the research sector, a higher level of data sharing may have a positive impact to the extent that such activities feed into the research process to increase the volume of research outputs of benefit to the wider society including regional communities. As discussed above, the findings from our commissioned report shows that research data provides a unique pathway to research impacts. Quality research data could be thought of as a public good. In the context of the work of ARDC, encouraging the sharing of research data is one way of maximising the return on public investment in research. The importance of data sharing has been clearly illustrated by the challenges posed by the Covid 19 international public health emergency. Rapid data sharing has been necessary for governments around the world to devise their public health responses, implement border control measures, and take other prompt actions to contain outbreaks within their jurisdictions. In addition, rapid data sharing has also made possible intensive scientific research into the virus and accelerated the development of vaccines, diagnostic tools and other therapies to treat patients infected with the virus.

# What are the greatest barriers to sharing data (either your own or other data) and how can these be overcome?

Limitations in the design of a data repository may deter data owners or custodians from depositing and sharing data assets with the general public. We will elaborate on this comment in the following section.

#### III. Differences between Open Data and FAIR Data

The consultation paper concludes with Appendix A titled, "Designing the Regional Data Hub." Appendix A begins by informing the reader that "the principles of open data and data exchange are central to the Hub (Figure A1)."<sup>14</sup> Through this initial statement, BCARR has indicated its intention to adopt a design approach guided by data principles. Yet, little detail is offered about what the chosen principles would mean for the construction and operation of the Hub. Instead of a discussion on open data and data exchange, the audience is referred to Figure A1 labelled, "Proposed Hub development." The content of Figure A1 does not appear to be consistent with common understandings of open data. The Open Data Handbooks offers the following definition.

Open data is data that can be freely used, reused and redistributed to anyone - subject only, at most, to the requirement to attribute and share alike.<sup>15</sup>

A similar perspective is adopted by the Open Knowledge Foundation: "Open means anyone can freely access, use, modify and share for any purpose (subject, at most, to requirements that preserve provenance and openness)."<sup>16</sup> Contrary to making data freely available for anyone to access, the ideas proposed in Figure A1 appear to suggest constraints and limitations in data sharing imposed either by (i) individuals and organisations with data holdings ("provide for data owners to set permissions") or (ii) governments and authorities exercising their regulatory powers ("deliver through a platform that complies with national data security and governance standards").<sup>17</sup> In the discussion on "data sharing opportunities", the consultation paper recognises that there are legitimate reasons to protect data and regulate access in order to "provide confidence and trust in data sharing" (e.g. privacy standards, confidentiality obligations, and cybersecurity regulations).<sup>18</sup> Due to the apparent tensions and inconsistencies in the proposed design of the hub, ARDC advises that BCARR review its choice of data principles for the design of the Regional Data Hub.

In place of open data, ARDC suggests that the FAIR data principles are probably more appropriate to the aspirations of the Regional Data Hub. Data is FAIR if it is "Findable," "Accessible under well defined conditions," "Interoperable" and "Reusable."<sup>19</sup> According to Mons et al, FAIR refers to

a set of principles, focused on ensuring that research objects are reusable, and actually will be reused, and so become as valuable as possible ... [These principles] describe characteristics and aspirations for systems and services to support the creation of valuable research outputs that could then be rigorously evaluated and extensively reused, with appropriate credit, to the benefit of both creator and user.<sup>20</sup>

<sup>&</sup>lt;sup>14</sup> Above n 1, 9.

<sup>&</sup>lt;sup>15</sup> Dietrich, Daniel, Jonathan Gray, Tim McNamara, Antti Poikola, P. Pollock, Julian Tait, and Ton Zijlstra. *Open Data Handbook* (2009, revised 2017) <https://opendatahandbook.org/guide/en/what-is-open-data/>.

<sup>&</sup>lt;sup>16</sup> Open Knowledge Foundation, *The Open Definition* (London, UK) <https://opendefinition.org/>.

<sup>&</sup>lt;sup>17</sup> Above n 1, 9.

<sup>&</sup>lt;sup>18</sup> Above n 1, 7.

<sup>&</sup>lt;sup>19</sup> Mons, Barend, Cameron Neylon, Jan Velterop, Michel Dumontier, Luiz Olavo Bonino da Silva Santos, and Mark D. Wilkinson. "Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud." *Information Services & Use* 37, no. 1 (2017): 49-56. <sup>20</sup> Ibid 50.

The FAIR Data principles were initially proposed in a 2016 paper authored by scholars working primarily in the life sciences.<sup>21</sup> Though its origins reside within a field of scientific inquiry, advocates argue that the FAIR principles may be "equally applied to any data, or any service, in any discipline."22 At the same time that they emphasise that "FAIR is not equal to open," Mons et al report that the FAIR principles were "inspired by Open Science."<sup>23</sup> The notion of the "FAIRness" of data can be thought of as a spectrum or "a continuum of increasing reusability" of data made possible through different modes of implementation.<sup>24</sup> The connotation of FAIRness associated with FAIR data implies appropriate or even-handed evaluation of competing claims to data or information. Facilitating or encouraging the publication of FAIR data is one means of progressing the goals of Open Science. Thus, the FAIR Data principles can be found in the latest draft text of the UNESCO Recommendation on Open Science. We note that in its 2021 draft Recommendation, UNESCO has described the FAIR principles as an example of the "principles of good data governance and stewardship."<sup>25</sup> Since its introduction, the FAIR Data principles have been adopted by governments, universities, and funding organisations. In Australia, these principles have been adopted in the Australian Government's National Collaborative Research Infrastructure Strategy.<sup>26</sup> In addition, the FAIR Data principles are also mentioned in the Open Access Policy (2018) issued by the National Health and Medical Research Council.<sup>27</sup>

#### Conclusion

The Regional Data Hub is a timely initiative which will serve a growing population of Australians living outside of the major capital cities. Through the *Regional Movers Index*, the Commonwealth Bank and the Regional Australia Institute have recently released data indicating the highest level of regional migration since 2018.<sup>28</sup> Their study found that net regional migration increased by 66 per cent. The recent appeal of regional Australia is in part linked to housing affordability, lifestyle choices, uptake in remote work, and growth in regional centres which is contributing to Australia's economic recovery.<sup>29</sup>

<sup>&</sup>lt;sup>21</sup> Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg et al. "The FAIR Guiding Principles for scientific data management and stewardship." *Scientific data* 3, no. 1 (2016): 1-9.

<sup>&</sup>lt;sup>22</sup> Mons et al, above n 19, 52.

<sup>&</sup>lt;sup>23</sup> Mons et al, above n 19, 51.

<sup>&</sup>lt;sup>24</sup> Mons et al, above n 19, 50.

<sup>&</sup>lt;sup>25</sup> See clause 7: UNESCO, Draft Text of the UNESCO Recommendation on Open Science (March 2021) <https://unesdoc.unesco.org/ark:/48223/pf0000376893?posInSet=7&queryId=64f6c09b-9508-4258-82a1-e195d9d38368>.

<sup>&</sup>lt;sup>26</sup> The FAIR Data Principles has been incorporated into the NCRIS Principles. Australian Government Department of Education and Training, *National Collaborative Research Infrastructure Strategy (NCRIS)* 2018 Guidelines <a href="https://www.dese.gov.au/national-research-infrastructure/resources/ncris-2018-guidelines">https://www.dese.gov.au/national-research-infrastructure/resources/ncris-2018guidelines</a>, see Appendix A.

<sup>&</sup>lt;sup>27</sup> National Health and Medical Research Council, *Open Access Policy* (November 2018) <https://www.nhmrc.gov.au/about-us/resources/open-access-policy#>.

<sup>&</sup>lt;sup>28</sup> Commonwealth Bank of Australia, *More Aussies choose regional living*, media release, 28 June 2021 <<a href="https://www.commbank.com.au/articles/newsroom/2021/06/aussies-choose-regional-living.html">https://www.commbank.com.au/articles/newsroom/2021/06/aussies-choose-regional-living.html</a>.
<sup>29</sup> Nathan Morris, "COVID-19 helping drive regional growth as flexible work sees people look beyond the city", *ABC Southern QLD*, 27 December 2020 <<a href="https://www.abc.net.au/news/2020-12-27/covid-19-drives-regional-growth/13009510">https://www.abc.net.au/news/2020-12-27/covid-19-drives-regional-growth/13009510</a>>; Nathan Morris, "Inland Queensland's top destination for city migrants and people fleeing coastal "yuppiedom", *ABC Southern QLD*, 28 June 2021

In this paper, we have sought to contribute to the public discussion on the Regional Data Hub by responding to selected questions and commenting on some issues presented in the consultation paper. To summarise our submission, we offer the following recommendations.

(i) That input from the research sector should be sought at the outset. Researchers and research organisations may be in a position to share their data assets and offer their expertise to contribute to the co-creation of the Hub. Equally, if researchers collaborate with regional communities to analyse, integrate and process complex data feeds, then more value will be unlocked for those communities. Researchers and research organisations should be seen as key stakeholders.

(ii) BCARR may want to consider approaching ARDC (regarding RDA), other NCRIS facilities, and the Indigenous Data Network to form partnerships to pursue mutual interests. These organisations could contribute their expertise and offer assistance on collecting and publishing specific types of regional data.

(iii) The FAIR Data Principles could offer a helpful resource to inform the design of the Regional Data Hub. (Re)defining the "A" in the FAIR acronym as "accessible under well defined conditions" introduces some flexibility to permit limited access where open data or unrestricted access is not possible.<sup>30</sup>

## About the ARDC

The ARDC is a transformational initiative that enables Australian research community and industry access to nationally significant, leading edge data intensive e-infrastructure, platforms, skills and collections of high-quality data.

The purpose of the ARDC is to provide Australian researchers with competitive advantage through data, providing access to leading edge eResearch collections, tools, infrastructure and services. Its mission is to accelerate research and innovation by driving excellence in the creation, analysis and retention of high-quality data assets.

<sup>&</sup>lt;https://www.abc.net.au/news/2021-06-28/inland-queensland-a-top-destination-for-city-migrants/100242006>.

<sup>&</sup>lt;sup>30</sup> Mons et al, above n 19, 51.