AUSTRALIA'S DATA-ENABLED RESEARCH FUTURE:

HEALTH AND MEDICAL SCIENCES

Australian Academy
of Health and
Medical Sciences

A collaboration between ARDC, ACOLA and
Australia's Five Learned Academies

Sponsored by

ARDC
Australian Research Data Commons

About this report

This project is the result of a partnership between the ARDC, Australia's five Learned Academies and ACOLA to ensure Australia can undertake excellent data-enabled research across all fields of research. Notably, the project sought to help build a more coherent data policy and strategic data planning environment to uplift national data infrastructure. Five domain reports were developed, and a synthesis report focused on common themes and multidisciplinary opportunities and needs. We hope that this project will transition into an ongoing national data policy and strategic planning capability.

Report prepared by Lanika Mylvaganam and Catherine Luckin.

This project received investment from the Australian Research Data Commons (ARDC). The ARDC is supported by the National Collaborative Research Infrastructure Strategy (NCRIS). We acknowledge both the funding and project support from the ARDC, especially from Dr Adrian Burton and Shannon Callaghan.

Date of publication:

June 2022

Publisher:

Australian Academy of Health and Medical Sciences

PO Box 6114

Woolloongabba QLD 4102

www.aahms.org

Suggested citation: Australian Academy of Health and Medical Sciences (2022) *Australia's Data-Enabled Research Future: Health and Medical Sciences.*

Acknowledgement of country

The Australian Academy of Health and Medical Sciences acknowledges the traditional custodians of the land on which our offices stand and on which we hold our meetings and events across the country. Aboriginal and/or Torres Strait Islander peoples were the nation's first scientists, and they remain the spiritual and cultural custodians of their land. We pay our respects to elders past and present.

About the Academy

The Australian Academy of Health and Medical Sciences is the impartial, authoritative, cross-sector voice of health and medical science in Australia. We advance health and medical research in Australia and its translation into benefits for all, by fostering leadership within our sector, providing expert advice to decision makers, and engaging patients and the public.

We are an independent, interdisciplinary body of Fellows – elected by their peers for their outstanding achievements and exceptional contributions to health and medical science in Australia. Collectively, they are a representative and independent voice, through which we engage with the community, industry, and governments.

The Academy is uniquely positioned to convene cross-sector stakeholders from across Australia to address the most pressing health challenges facing society. We focus on the development of future generations of health and medical researchers, on providing independent advice to government and others on, and on providing a forum for discussion on progress in health and medical research with an emphasis on translation of research into practice.

The Academy is registered with the Australian Charities and Not-for-profits Commission (ACNC) and is endorsed as a deductible gift recipient.

ABN: 55 167 124 067

www.aahms.org

Table of contents

Executive summary

Health data have for decades been an important tool in better understanding disease, providing better care and treatment, and improving health. The recent exponential growth in the collection and availability of electronic health data presents new opportunities to advance health and medical research. The infrastructure, assets, policies and skills that enable the collection and effective use of these data have been progressing. However, health and medical researchers still report barriers preventing many of these opportunities from being realised.

To capitalise on the opportunities to improve health and make use of the ever-increasing volume of health data, Australia must develop an environment that enables the safe and secure use of patient data for legitimate research purposes – balanced with protecting the rights and interests of individuals. Collaboration and coordination within and across research disciplines will be key to achieving this.

To help advance Australia's research data infrastructure, assets, policies and skills, the Australian Academy of Health and Medical Sciences (AAHMS) has partnered with the Australian Research Data Commons (ARDC), the other Learned Academies (Australian Academy of Science; Australian Academy of Engineering and Technology; Australian Academy of the Humanities; Academy of the Social Sciences in Australia) and the Australian Council of Learned Academies (ACOLA), to better understand the current and future data needs and requirements of researchers across a range of disciplines.

The focus of the AAHMS contribution to this project was to seek insights from across Australia's health and medical research landscape, and to hear directly from researchers about their experiences using data to achieve their research goals. This report details the results of an environmental scan of the data landscape, as voiced by these health and medical researchers and through desktop research. The report discusses key challenges associated with data-enabled research, explores gaps in data infrastructure, assets, skills and policies, and identifies opportunities for Australia to enhance its management of health and medical research data infrastructure.

The results of this environmental scan will contribute to a consolidated report that brings together findings from all Australia's Learned Academies. This project is intended to seed a long-term partnership between ARDC, ACOLA, the Learned Academies and their Fellows to help advance strategic capability planning for national research data infrastructure.

Key findings

More coordinated and coherent data infrastructure, assets, policies, governance and ethics processes would enable better health and medical research. Researchers struggle to navigate this diverse landscape due to the fragmentation and inconsistency that exists, particularly across the Commonwealth and the various states and territories. Streamlining and harmonising these systems, policies and processes would help mitigate the challenges facing health and medical researchers.

Developing a national linked data asset for use across disciplines and sectors that follows the FAIR (findable, accessible, interoperable and reusable) and CARE (collective benefit, authority to control, responsibility and ethics) principles would create opportunities for research and would better facilitate research using linked person-level data. A number of important linked data assets are in use across Australia today. However, there is a lack of cohesion between assets, there are data gaps within assets, and there are very significant challenges associated with accessing linked data assets. A non-disease specific “plug and play” national linked data spine could address these issues.

Growing and nurturing data skills in research and within the health system will be crucial to advancing the future of data-intensive health and medical research in Australia. Investment in strategic workforce planning and training to improve data skills, including those for data collection, analysis, interpretation, communication and development of data-informed policy is needed.

It is crucial that the research community and data custodians build a social licence for data use in health and medical research – by growing public awareness and trust, seeking to understand public perceptions and concerns, and protecting the privacy of data collected from individuals. This would enhance the capacity of researchers to undertake data-intensive research that can improve patient outcomes and the health of the nation. Ongoing assessment of data privacy, public awareness and levels of trust will help gauge the level of social licence over time, and provide a mechanism to better understand the impact of these factors on research.

1. Introduction and context

Background

The health data landscape in Australia has undergone profound changes over the past few decades. Increasing digitisation of health information has paved the way for new methods of producing, collecting, storing, and analysing data.[1] Emerging disciplines like health data science have broadened understanding of real-world health problems and big data is driving new approaches to health and healthcare in Australia and globally. One study estimated that as much as 30% of the world's total stored data are generated in the healthcare industry alone and further research suggests the amount of global health data being generated is growing by 48% per year.[2,3]

What are health data?[5]

Health data relate to information gathered on people's health – and factors that affect it – and the health system. These data can be collected from a wide range of sources and include, for example, patient health records, blood or tissue samples, imaging data and health/fitness trackers.

Health and medical research has been a major beneficiary of these changes, with a noticeable shift in research patterns and methodologies in recent years.[4] The rapid growth of digital technology has meant data can be collected from a diverse range of sources including electronic medical records from hospitals and primary care, administrative data sources (such as Medicare claims), clinical trials and other types of health and medical research, wearable health tracking devices and more.[5] There are a growing number of examples from Australia and around the world that demonstrate how the use of big data can make a quantifiable contribution to population health and health policy with the potential to significantly impact patient outcomes.[6–10] Researchers today have more opportunities to utilise these health data to answer a greater breadth and depth of research questions to benefit the health of Australians and the global community. Through research, the use of health data can fundamentally shift the needle on Australia's most pressing health challenges. COVID-19 provides a stark example of this – highlighting both the crucial role data have played in the global response to the pandemic and the stark knowledge gaps resulting from a lack of data. Governments have relied on data across multiple sectors to make timely, reliable, evidence-based decisions that have directly impacted population health. [1]

In addition to vast opportunities, the exponential growth of health data and the evolution towards more data-enabled research have presented a number of new challenges that have impacted progress towards the most efficient and effective use of data in health and medical research. The 'FAIR Guiding Principles for scientific data management and stewardship' were published in 2016.[11] They were developed as guidelines to improve the findability, accessibility, interoperability and reuse (FAIR) of digital assets. While there has been progress in building the necessary data infrastructure, fully findable, accessible, interoperable and reusable data resources are still uncommon.[12] Linking data from different sources is another example of a challenge that, if overcome, would better enable researchers to gain a more holistic and longitudinal understanding of representative population groups with less data bias, while potentially reducing data collection costs.[13] Fundamental questions that challenge assumptions around historical, current and future Indigenous health data ownership, representation and control are also being asked both in Australia and around the world.[14] Data that pertain to Aboriginal and Torres Strait Islander peoples and their use are seen by many First Nations researchers as being obstructive of Indigenous goals rather than assisting their progress.[15]

The Australian healthcare system, like many others globally, is facing considerable challenges that if left unresolved, could threaten the health of all Australians. An ageing population, climate change and the emergence of new infectious diseases are just a few of the complex issues that will have an impact. Optimal health data infrastructure, assets, policies, and skills for use in research and across the health system will be essential elements in tackling these challenges now and into the future.

This report outlines the findings of our environmental scan. Chapter 2 outlines the project approach. Chapter 3 provides background on the scenarios used as a basis for developing the environmental scan, which are attached in full in appendixes C, D and E. These scenarios are then highlighted as examples in boxes throughout the report. Chapter 4 provides context on Australia's current health and medical research data landscape. Chapter 5 outlines our findings on what Australian researchers' need and require from national data infrastructure, assets, policies, and skills, which would best assist them in advancing the nation's health through research and achieving the positive outcomes outlined in the scenarios. Chapter 6 describes the key challenges faced by researchers as they undertake health research using data. Chapter 7 identifies gaps and opportunities, which if addressed could help advance and optimise data-enabled health and medical research in Australia. Chapter 8 considers the role of AAHMS in supporting these endeavours.

2. Project outline and methodology

Health and Medical Data Steering Committee

The development of this environmental scan has been led by the AAHMS Health and Medical Data Steering Committee comprising of seven interdisciplinary Academy Fellows with expertise in data-enabled research. The steering committee met virtually between October 2021 and February 2022 and was chaired by Professor Louisa Jorm FAHMS.

Health and Medical Data Steering Committee

- Professor Louisa Jorm FAHMS (chair)
- Professor Melanie Bahlo FAHMS
- Professor Julie Byles AO FAHMS
- Professor John Carlin FAHMS
- Professor Annette Dobson AM FAHMS
- Professor Stephen Jan FAHMS
- Professor Anton Peleg FAHMS

Appendix A provides further detail on the expertise and experience of the steering committee members.

Project approach

The methodology for this project was developed in collaboration with ARDC, ACOLA and the other Learned Academies.

Scenario planning

To test the current and future data-enabled research needs and requirements within the health and medical research landscape, AAHMS developed three scenarios to facilitate discussion of a range of challenges facing the sector. Scenario planning and discussion was the main tool used by all Academies for data gathering in this project.

In developing the scenarios, the AAHMS steering committee considered a variety of issues that currently impact on Australia's health and are predicted to do so in the future. Australians face a growing number of important health challenges. The steering committee sought to identify a set of scenarios for this project that collectively represent a range of complex, multifactorial issues, impacting people from diverse backgrounds and requiring cross-sector solutions. The next section of this report provides an overview of the scenarios, which focussed on:

- Dementia
- Obesity
- Antimicrobial Resistance

Consultation

In December 2021, AAHMS convened one roundtable for each of the three scenarios which were attended by more than 50 contributors in total, representing diverse disciplines and sectors. Participants were asked to identify opportunities, challenges, gaps, and priorities within the scenarios associated with Australia's data infrastructure, assets, policies, and skills.

The full scenarios presented to roundtable participants can be found in Appendices C–E (available in separate document).

Desktop research

In addition to scenario planning, the project included desktop research to identify existing research priorities, data assets, policies, projects, stakeholders and skills in Australia to guide the development of this environmental scan.

Exposure draft

An exposure draft of the AAHMS environmental scan was circulated to all roundtable participants and other stakeholders for comment – with feedback incorporated in liaison with the steering committee.

This project does not provide a comprehensive report on the data landscape of Australia's research environment. As such, it is not intended to describe:

- a full inventory of all applications, data sets or systems desired by Australian researchers
- a comprehensive picture of the international landscape, or
- the budgeting, design and implementation plans of any specific data capability solution.

3. Overview of scenarios

The scenarios used in this project were selected and designed to outline the current status of different health challenges and to frame the discussion of those challenges in a positive vision for the future. In this future, the associated health impacts have been mitigated, the health of the population has been improved and Australia's health system has not been overwhelmed.

Status quo of dementia, obesity and antimicrobial resistance in Australia today

Dementia

Dementia describes a collection of symptoms that are caused by disorders affecting the brain. Dementia affects thinking, behaviour and the ability to perform everyday tasks.[16] Australia's population is ageing, with the number of Australians aged 65-69 years expected to increase by about one-third, and those aged 85 and over predicted to double over the next three decades.[17] In 2018, dementia was the second-leading cause of death in Australia, and the most recent report from the Australian Institute of Health and Welfare (AIHW) estimates that between 400,000 and 460,000 Australians have dementia.[17] Although dementia can affect younger people, it is more common with advancing age, leaving the growing population of older Australians at risk of significant levels of disability and dependency. Consequently, this also places an ever-increasing burden on the rest of society, with growing demands for health and social care. Dementia cost the Australian economy approximately $15 billion in 2018. This is predicted to rise to more than $36.8 billion per year by 2056 without further intervention.[18]

Obesity

Overweight and obesity refers to excess body weight. Body mass index (BMI) is an internationally recognised standard for classifying overweight and obesity in adults.[19] The most recent report from the AIHW shows that in 2017-18, an estimated two in three (12.5 million) adults in Australia aged 18 years and over were affected by overweight or obesity.[20] Alarmingly, one in four (1.2 million) children and adolescents aged 2–17 years were categorised as overweight or obese, increasing their risk of poor physical health, illness and early mortality in adulthood.[21] Obesity and overweight is influenced by a range of factors including food and nutrition, social norms, physical activity, and other biological and environmental factors that have their impact across an entire lifetime. Obesity and overweight is a risk factor for many of Australia's most prevalent diseases, from coronary heart disease to diabetes, some cancers and more.[22] It is also a likely risk factor for dementia (via coronary heart disease). Because of this, it is one of five key risk factors that are associated with the greatest burden of disease with one study estimating that by 2060, it could cost the Australian economy approximately $140 billion per year.[23]

Antimicrobial resistance (AMR)

Antimicrobial resistance (AMR) occurs when bacteria, viruses and other pathogens evolve over time so that antimicrobial medicines become ineffective, and infections become increasingly difficult or impossible to treat. If left unchecked, AMR has the potential to undermine major medical advances with devastating impacts. On the current trajectory, AMR will significantly limit the effectiveness of lifesaving treatments used to fight infections such as pneumonia, salmonellosis, gonorrhoea, and infections resulting from chemotherapy and surgeries such as organ transplants, caesarean sections, and care of pre-term babies.[24,25] By 2050, it is estimated that AMR could result in over 10,430 deaths annually in Australia. In addition, one model suggests that by 2050, the far-reaching impacts of AMR on the treatment of infections could cost the Australian economy between $142 billion and $238 billion annually.[26] The impact of AMR on human health is only one piece of the puzzle, with animal and environmental health simultaneously at risk.[27]

A positive future

Each scenario asked researchers to imagine a future from 2050 onwards where today's projections for the impacts of dementia, obesity and AMR have not been realised because of interventions taken to address them. For instance:

- Risk factors for dementia have been reduced thereby improving disease prevention, and for those diagnosed with the condition, better detection, reporting, management and support protocols have reduced individual, societal and economic burden.
- Obesity rates have declined, Australians are more active, have improved nutrition and lead overall healthier lives, all of which have contributed to the prevention of chronic diseases and mental illness. Priority populations once most likely to experience obesity, now have rates similar to the general population.
- Existing medicines remain effective against infections while development of new antimicrobials is occurring at a rate sufficient to curb the spread of AMR. Antimicrobial misuse in humans, animals and plants has declined and safer usage practices continue to be implemented.

The aim of setting out the future scenarios in this way was to encourage contributors to consider how data can enable research to make advances that mitigate risks and bring about these more positive futures. The scenarios anticipate that advances across a diverse range of areas facilitate many of these outcomes. Cross-sector and interdisciplinary collaboration, the discovery of new technologies, a focus on the social determinants of health, sustained political will, funding and coordination across systems are just a few examples of key influencing factors. Through the roundtables, contributors were asked their perspectives on the data needs and requirements that would help them support and deliver these advances through their research. They were asked to consider these (and other) factors, the dynamics of how they interact and on what timeline. The scenarios encouraged researchers to go beyond the status quo and imagine novel, impactful solutions – particularly those enabling collaboration across domains – while considering the current and future opportunities, challenges, gaps and priorities for Australia's data landscape.

The outcomes of these discussions have informed this report.

4. Australia's current health and medical research data landscape

The current health and medical research data landscape is vast, with a complex network of data assets, policies, skills and stakeholders. It was not within the scope of this project to provide a comprehensive map of the landscape, however, some examples of existing data infrastructure, assets, polices, and skills, can be seen below.

Infrastructure and assets

There is an extensive network of existing infrastructure and assets that serve various purposes for health and medical research in Australia. For example, the Medical Benefits Schedule (MBS) data collection contains information on services that qualify for a benefit under the *Health Insurance Act 1973* and for which a claim has been processed. The database comprises information about MBS claims (including benefits paid), patients and service providers.[28] In addition, state and territory governments collect a range of data from different sources such as hospital data, registry data and patient and population surveys. There are also existing centres that provide data linkage services and infrastructure to Australian researchers. These include the AIHW, The WA Data Linkage System (WADLS), The Centre for Health Record Linkage (CHeReL), The Centre for Victorian Data Linkage (CVDL), the Tasmanian Data Linkage Unit (TDLU), SA NT DataLink and Data Linkage Queensland (DLQ).[29] Australia also has a number of national data infrastructure and assets currently under development such as My Health Record (MHR). More examples of Australia's data infrastructure and assets can be found in Appendix B.

Unfortunately, this fragmented data ecosystem creates consistent and persistent barriers to streamlined research. The data infrastructure and assets may not be easily available to researchers outside government agencies. In addition, data resources differ across jurisdictions, have various data owners and custodians, and inconsistent governance structures. Throughout the AAHMS roundtable discussions, it was clear that there is varied understanding of these resources, what they realistically provide and how to access them. It is also difficult to quantify how each of these resources contribute to health and medical research, with limited standardised metrics to measure the true utility of each resource.

National Research Infrastructure Roadmap

The current National Research Infrastructure (NRI) Roadmap identifies priority research infrastructure for Australia. Digital data, data infrastructure and integrated systems were specified as a focus area, highlighting the increasing dependence on data as a driver for change.[30] Despite advances in infrastructure for data-intensive research, health and medical research has not yet seen the hoped-for degree of improvement in the day-to-day processes of research, with much national data infrastructure designed mainly to support research that does not use potentially sensitive human data.

Progress has been made over the past decade. And although the NRI roadmap does not encompass the full range of issues facing health and medical researchers, it is heading in the right direction.

Policies

The policy and legislative environment for research data is complex and regulated by both Commonwealth and State and Territory legislation. In reality, these environments are therefore inconsistent, fragmented and unenforceable.[31] In addition to state and territory differences, there are inconsistencies between the public and private sectors. There is no single body with the remit to oversee the research data agenda.

Supporting Aboriginal and Torres Strait Islander data sovereignty and governance

The Maiam nayri Wingara Aboriginal and Torres Strait Islander Data Sovereignty Collective was formed in 2017 to develop Aboriginal and Torres Strait Islander data sovereignty principles, and to identify Aboriginal and Torres Strait Islander data assets.[32] This group aims to progress Aboriginal and Torres Strait Islander sovereignty and data governance through developing shared understandings and initiating an Australian set of Aboriginal and Torres Strait Islander Data Governance protocols.[32] The Collective continues to inform health data policy through its publications and networks.

Skills

Data and digital skills are among the fastest growing emerging skills in Australia. A number of world class biostatistics, bioinformatics and health data science education and training programs have, and continue to, produce high quality health and medical researchers. However, as the skills required for data-enabled research broaden, the knowledge gaps become more apparent.[33] Little progress has been made in developing a coordinated approach to data skills in the existing health and medical research workforce, and resources to train a new cohort of skilled health data science researchers have not been sufficient to match the increasing demand for these skills.[34] There remains limited systematic investments in the fundamental core disciplines that are needed to provide leadership in health data science, to ensure the greater availability of data is matched with a capacity to use those data sensibly and efficiently.[35]

Ultimately, the current capacity of Australia's health and medical data research infrastructure, policies and skills is not sufficient to realise the significant potential of a data-enabled research landscape.

5. Needs and requirements

During our evidence collection, we asked researchers from a range of backgrounds about their perspectives on what national data infrastructure, assets, policies and skills would best assist them in advancing the health of Australians through research, and achieving the positive outcomes outlined in each scenario framework.

Although there have been several important developments in Australia's health and medical data landscape, particularly in recent years, researchers report that they are yet to experience the full impact of these changes. In addition, advances in health and medical data infrastructure, assets, policies and skills have not appeared to keep pace with the rapid growth of health data and data-driven research. This is particularly evident for research investigating complex health issues that require interdisciplinary and cross-sector data solutions.

The Academy's data definitions

National data infrastructure – Nationally significant data assets, facilities, and services which are scalable and collectively managed and operated for use by research institutions, service providers and other public and private entities across the country.

Data assets – Nationally strategic databases, data systems, web-based resources or data services.

Data policies – Principles that describe the rules to control the integrity, security, quality and use of data at a national level.

Data skills – Expertise, experience, and qualifications to design studies, analyse, manage, store and distribute data.

Genomics data is one example of a rapidly evolving area associated with a number of complexities. Health genomics is increasingly being recognised as an area that could transform the way healthcare is delivered in Australia. Sequencing one person's genome can generate a huge amount of data that can be useful in other areas of health and medical research. To manage data of the scale and complexity associated with genomics data, scalable, secure, and interoperable analytics, data storage, infrastructure, and local and national services for data access and exchange will be required.[36] Furthermore, streamlined mechanisms to connect and integrate the outputs of genomic analyses to person-level phenomics and outcomes data generated by the health system is required to fully realise the potential benefits for health.

The following priority areas represent a range of the key data needs and requirements facing health and medical researchers today. The FAIR and CARE principles, linked data, the right characteristics of data, and data skills are areas that are still developing in Australia. Improvement here would help researchers make much better use of Australia's data landscape.

FAIR and CARE data principles

FAIR

The FAIR guiding principles for research data management and stewardship (findable, accessible, interoperable and reusable), were officially published in 2016.[37] Since then, they have been universally accepted as fundamental to good data standards and policies. In Australia, the adoption of the FAIR principles in health and medical research has been advancing in some areas, while in others there has been little progress. Importantly, the National Health and Medical Research Council (NHMRC) has specified its support for the FAIR principles within its Open Access policy, and key data leaders such as ARDC continue to encourage the use of FAIR data though strategic projects and partnerships.[38,39] However, there is still no coherent national research data policy which promotes the FAIR principles.[31]

Despite an abundance of health data, researchers experience myriad barriers to accessing data usable for research. Complexities within these non-standardised environments can cause significant impacts to research timelines, funding and methodologies. A 2018 report published by the Digital Health Cooperative Research Centre described a fragmented health research landscape that produces data silos, making data management ad hoc, inconsistent and unclear.[40] These characteristics are still being experienced by researchers today. Policies and systems that enable simple, efficient, dynamic and affordable access to, and use of, interoperable data are essential to advancing data-enabled health and medical research. Examples of the data infrastructure that make up the current fragmented landscape are presented in Appendix B.

CARE

Historically, data practices, such as the collection and use of data, have not served the interests of Indigenous peoples worldwide. Aboriginal and Torres Strait Islander researchers have raised concerns over the rights of Australia's First Nations peoples to data that involves their health, communities, cultures, languages and more. It is essential that policies and principles reflect these rights by ensuring that data is used for Aboriginal and Torres Strait Islander self-determined interests. The FAIR principles alone do not sufficiently acknowledge and address the power differentials and historic contexts around data sharing and data use for health and medical research.[41] Aboriginal and Torres Strait Islander researchers have contributed to the CARE principles for Indigenous Data Governance (collective benefit, authority to control, responsibility and ethics), which aim to complement the FAIR principles to ensure a more strategic and holistic approach to addressing Indigenous data ownership, sovereignty and governance.[42]

Data linkage

Data linkage is the process of bringing information together from different sources that relate to the same person or entities to create a new, richer data set.[43] With growing amounts of health data from a greater number of sources, data linkage is an invaluable tool for health and medical researchers. The AIHW is a national resource for a range of linked administrative data collections, and each state has its own centre for data linkage (South Australia and the Northern Territory have a joint centre).[29] The Australian Government has invested in developing Australia's data linkage capability through several mechanisms including the Population Health Research Network (PHRN). This network aims to support world-class research through privacy-preserving linkage of Australia's population-based health and other human services data.[44]

Despite this investment, health researchers continue to face very significant obstacles and inefficiencies, and the need for linked data is still growing. The obesity roundtable discussion highlighted the complexity of the current data linkage infrastructure (see Box 1). There is duplication and a lack of cohesion, the consequences of which are that researchers often struggle to conduct studies that require national record linkage in a timely manner.[45] Researchers indicated that in many cases they know the data are there, but accessing the linked data necessary for conducting their research is challenging due to the need for approvals from multiple gatekeepers. Another challenge is navigating the data governance, privacy and ethics implications of existing data linkages involving major administrative data sets. One researcher described the process of accessing data to be "like going around in a circle with agencies not speaking to each other". In addition, there are no national data sets that sufficiently address the health and lifestyle risk factors of many health challenges, the experiences of those that suffer them, their families and communities, or the cost to individuals and society. Researchers also expressed a requirement to link data that are generated through research (e.g. clinical trials, cohort studies) with routinely collected data in order to efficiently obtain information about long-term outcomes.

Box 1: Scenario case study – **Cross-sector data linkage needs for obesity research**

Obesity is a public health issue influenced by multiple sectors and industries. Food and nutrition, physical activity, built environments, transportation, social media, social norms, education and socioeconomic status are just some of the factors considered by researchers when examining this national health priority. Up to now, obesity has been underrepresented in research publications involving the use of linked data.[44] Access to many of the routinely collected data sources across these areas could significantly help researchers in their analyses of these complex and multifactorial influences. For example, research analysing the links between obesity and transport patterns is currently being undertaken. However, the current data infrastructure does not facilitate the linkage required for this type of research, making it very challenging to piece together the required information.

The recently released National Obesity Strategy 2022-2032 highlights the use of evidence and data as being one of three key enablers to guide implementation of the strategy.[75] Strengthening data infrastructure, and removing the barriers noted above, will be essential to ensure the successful implementation of this strategy.

Data types and characteristics

In conjunction with data that are FAIR and linked, researchers need access to data sets with the right kinds of high-quality data. The production, collection and input of unbiased, accurate and consistent data is fundamental to advancing the work of health and medical researchers. In addition to robust data sets, comprehensive data that include a wide range of health information are essential. Administrative or other data that are routinely collected are critical for data-driven research in the health and medical science. However, administrative data are often limited, with insufficient information on data quality and key confounders and risk factors such as body weight, height, socioeconomic status or smoking status.[34] The limitations in the use of observational data have also been recognised. Researchers also described the need for access to richer, "organic" data that are generated without any explicit research design elements and are continuously documented by digital devices (e.g. physiologic measurements that are captured by sensors and wearables, or content and social interactions extracted from social networking sites). Although these data are being collected at higher rates than ever before due to increased use of technology, including electronic medical records, there is still a way to go before researchers can readily access the information.

In the AAHMS discussion on dementia, the lack of data created a number of core issues for researchers (see Box 2). Contributors agreed that data capture at or around the time of diagnosis in primary care, community aged care or residential care settings would be hugely beneficial. Standardised collection of minimum data sets across these sectors is needed. However, these solutions alone will not address all the issues.

Box 2: Scenario case study – **Data for dementia research in Australia**

At present, national dementia data are limited. Unlike diseases such as cancer or diabetes, there is no system in Australia for monitoring dementia in the community.[76] Despite recent changes in the health and aged care policy and data environments, persistent gaps continue to hinder researchers. Key messages from the Academy's evidence collection for this report indicate that the current challenges facing dementia researchers span the breadth of the dementia data landscape. In addition, aged care data is seen as an example of how Australia has not only failed to progress in the provision of meaningful or useful data, but has in fact gone backwards. For example, the exact number of Australians with dementia is still unknown, and with limited data from diagnosis to management in primary, hospital, aged and residential care settings, researchers are lacking robust information to examine the true picture of dementia and dementia care in Australia.

Throughout the Academy's discussions, it was noted that changes to the dementia data landscape appear to focus on the purpose of funding services, with less consideration for evidence, equity, quality, timeliness and impact of those services. The gaps are exacerbated by different jurisdictions including health/state, aged care/federal. Contributors indicated the need for data that meets the FAIR and CARE principles, and that they require routine and comprehensive collection of broad-ranging data – including diagnostics, therapeutics, epidemiological life course, genomics, data on carers and more. An upskilled workforce across the dementia pipeline with training in the value of data and data literacy was also expressed as a highly sought-after change.

Data skills and expertise

The rise of data-enabled research in health has exposed a shortage of accompanying data skills within Australia – including those for data collection, analysis, interpretation and development of data-informed policy. This has hindered research groups in their efforts to build interdisciplinary teams with the necessary data skills to enhance their approaches to this type of work. Many teams are forced to look outside Australia for these skills, and with the COVID-19 pandemic leading to border closures, existing workforce shortages have been exacerbated and brought to the fore.

Some of the key data skills researchers would like to include in their research groups include:

- Data/analysis communication skills
- Ability to work across traditional discipline silos
- Ability to integrate different stages of the data pipeline
- Understanding of both qualitative and quantitative data
- Cross-sector data understanding

- Ability to understand and address cultural sensitivities within data collection, use, analysis and interpretation
- Skills in using data to inform policy and communicating with policymakers
- Skills in dynamic systems modelling
- Skills in data science, including machine learning and artificial intelligence (AI)
- Deep understanding of how to frame answerable research questions (descriptive, predictive, causal) and design studies using appropriate data methodologies
- Skills in modern methods that prioritise causal thinking for answering questions about interventions (actual or potential) for improving health

By investing in strategic workforce planning and proper training, Australia can reap long-term and far-reaching benefits in the research sector and beyond. Building an environment that nurtures, sustains and grows this workforce will be critical. The education sector can play an important role here by providing a pool of trained graduates with both technical skills and a broader understanding of ethics, privacy and safety. Another important part of building an Australian workforce will be to ensure we can also attract international talent to enhance our expertise and experience. Furthermore, ensuring that there are incentives for data scientists to seek a career in health and medical research, noting strong demand from both public and private sector organisations that offer competitive benefits and greater job security, will be crucial.

We also heard from researchers about the need to build data skills across primary, secondary, hospital, residential and aged care settings. Proper data collection, entry and stewardship within these settings is essential to building quality data for use further down the data pipeline. With the right infrastructure, including software, improving data literacy and skills within these workforces should help achieve this.

Social licence for the use of big data

The concept of "social licence" refers to the informal permissions granted to organisations and institutions by members of the public to carry out a particular set of activities.[46] Social licence is intrinsically linked to public trust and where the activities of organisations and institutions are trustworthy, they are more likely to have social licence to engage in these activities – such as the collection and use of health data. Social licence is complex, challenging to quantify and there is limited literature that usefully measures the full scope of public trust in data for use in health and medical research. Despite this, it is generally accepted that the benefits of using the concept of social licence as part of a strategy to scale up data-enabled research outweigh any perceived or actual risks.[47]

Expanding public discourse to build greater public awareness, understanding and trust around the benefits of health data use in research to inform policies and programs and improve patient outcomes could considerably grow the capacity of researchers to undertake data-enabled research. This needs to be informed by work to understand public perceptions of the opportunities and concerns they may have in relation to the use of health data for research purposes. Ultimately, Australia needs to create an environment that enables the safe, secure and transparent use of patient data for legitimate research purposes, while also protecting the rights and interests of individuals.

6. Challenges

Inevitably, the rapid growth and use of data in health and medical research has presented a number of challenges. As the volume of health data generated continues to expand exponentially, and more attempts to conduct data-enabled research are made, the greater the impact on researchers, the research they conduct and the public they seek to benefit.

In the roundtable discussions, researchers articulated a number of significant challenges that are echoed in current literature on data use in health and medical research. In addition to the earlier issues, raised in association with Australia's health and medical research data needs and requirements, the following areas outline some of the more pressing challenges requiring priority action.

Governance, ethics and privacy

Complex governance and ethics approvals processes have acted as a barrier that has been embedded in the health and medical research ecosystem for many years. These processes differ across the full range of data custodians including organisations, institutions, and state and federal systems.[40] An example of this is the complex ethics and regulatory frameworks around data stored in Australia's biobanks, which are controlled at a state and territory level and are not easily integrated into national systems. The lack of consistency and connectivity across processes makes it challenging for researchers to know what to expect and plan for. This can have a significant impact on project timelines and funding, adding unnecessary complexities for researchers to manage with very little support. In addition, researchers indicate a lack of transparency around these frameworks which exacerbates existing barriers. The 2021 NRI roadmap exposure draft (see Appendix B) responds to some of these concerns by highlighting the need for streamlined governance structures and recommending an expert NRI advisory group to drive a more effective ecosystem.[48] However, some researchers have expressed doubts that this approach goes far enough to ensure the fundamental challenges are being addressed with sufficient urgency to match the ever-increasing impacts.

Concerns around the privacy of health data are very high and as sensitive data continue to be collected and stored at unprecedented rates, efforts to resolve critical privacy-related issues continue to face a number of challenges. There is ongoing debate as to whether the privacy interests of individuals can appropriately align with the public good. At present, there does not appear to be a straightforward solution to this challenge.[49] However, there is growing alignment in Australia around use of the "Five safes" framework (safe people, safe projects, safe settings, safe data, safe outputs), which presents a holistic approach to considering and managing the risks of data sharing or release.[50]

As highlighted in the previous section of this report, considerations around health data protection predominantly revolve around issues of public trust. Research Australia's most recent public opinion polling suggests that the use of de-identified medical records for research is strongly supported by 35.8% and somewhat supported by 47.0% of people. The strongest support comes from Australians aged 65 and over, which raises potential future questions around the evolving landscape of data privacy considerations for younger cohorts.[51] Some of the other moving parts that will influence issues around data privacy include:[49]

- Data protection laws
- Informed consent
- Re-identification of anonymised data
- Nefarious use of data
- Opt-in versus opt-out models of consent

Achieving and maintaining appropriate levels of information security, confidentiality and privacy will continue to pose unique challenges to the health and medical research sector. Approaches to resolving these challenges will require strategic consideration of both the current and long-term implications for individuals and communities.[52]

Data bias

There are differing understandings of bias in data across various areas of research. Examples of problems include:

- Under-reporting and poor data quality, such as for Aboriginal and Torres Strait Islander status and languages spoken.
- Under-representation of certain groups in some data collections, such as people who may not have access to digital devices.
- Inappropriate data sets being used to train machine learning/AI algorithms, such that they perpetuate biases in past or current practices.

Despite being a relatively new field, health data science has unsurprisingly been shaped by the social forces impacting the individuals and communities that the data are built from. Aboriginal and Torres Strait Islander peoples have experienced racism within the health sector for decades, and consequently interact differently with the health system compared to other members of the Australian population.[53] Many Aboriginal and Torres Strait Islander peoples describe the values they hold about health and wellbeing as not being well understood or reflected in the care they receive. This in turn has an impact on the care they seek.[53] Culturally and linguistically diverse (CALD) communities also face multiple challenges accessing and utilising health services.[54] Similarly, patients living in rural or remote areas in Australia may be under-represented in some data sources, either because they lack access to services, utilise other services funded under different arrangements, or because different data collection arrangements apply. In this way, biases are easily encoded in data sets and in the application of data science, with the potential to reinforce existing social injustices and health inequalities.[55]

The use of data collected by digital devices such as smartwatches, fitness trackers and smartphones are also increasing. There may be a level of unconscious bias embedded in these devices that requires further investigation. Early studies have noted a specific risk of systematically undercounting of steps for people with obesity, females or people from different ethnic groups [56]. In addition, the demographics of people likely to own these devices may also contribute to bias, depending on how the data are used. Nonetheless these devices offer significant research opportunities.

Data sharing

Over many years, the health and medical research community has expressed a particular concern relating to the availability and transparency of both government (public) and private sector data for research. Both of these areas are hold rich and potentially significant data that researchers are not able to tap into efficiently or effectively. The Data Availability and Transparency (Consequential Amendments) Bill 2020 is currently before the House of Representatives. If passed, this bill will implement a scheme to authorise and regulate access to Australian Government data.[57] Private sector data is also critical to a comprehensive analysis of Australia's health landscape. Some of the highest value private data collections that could benefit health and medical research include:

- Private hospital data
- Health insurance data
- Pathology
- Radiology
- Dental
- Health and wellbeing data from other sources including wearable devices
- Food purchases
- Patterns of transport
- Social services
- Education
- Criminal justice

There are a number of challenges to consider when examining the role of data sharing for health and medical research. Questions around data privacy and security require serious scrutiny, as noted above. In addition, many of the complexities around data sharing in Australia stem from a culture in which stakeholders do not always support or promote openness and

Box 3: Scenario case study – **The challenge of developing a data-driven research agenda for antimicrobial resistance (AMR)**

AMR is a prime example of a large scale, cross-sector health issue that could benefit from a coherent data-enabled research strategy. However, the Academy's evidence collection for this project highlighted the complexities around attempting to define a data-driven research agenda when the core problems are currently not well understood. Researchers broadly agreed that while data will be essential to tackling the issue, the more pressing challenges in AMR research lie with first identifying the research questions that will have the biggest impact.

Tackling AMR requires One Health solutions. One Health is a collaborative, multisectoral, and transdisciplinary approach – working at the local, regional, and global levels – with the goal of achieving optimal health outcomes that recognise the interconnection between people, animals, and their shared environment. Research that prioritises One Health is growing however, there is still a lot to learn in this space and silos remain between the sectors, limiting the capacity of available data to have its greatest impact on AMR. There are a number of ongoing research projects and programs across health, animal science and environmental science organisations with differing levels of coordination between and across sectors. In addition, data infrastructure, governance and policies vary across sectors, and there is little harmonisation between states and territories.

transparency. Shifting this paradigm will be a long-term endeavour requiring the necessary level of incentivisation and buy-in across all sectors. It was not within the scope of this project to critically assess the benefits and pitfalls of data sharing or to analyse when data sharing would be most appropriate. However, regardless of the associated challenges, it is important to be aware that data sharing has the potential to bring significant benefits to the health and medical research sector.

7. Gaps and opportunities

Throughout this report, we have identified numerous areas that can be advanced for the benefit of data-enabled health and medical research in Australia. An optimised data landscape has considerable untapped potential to influence Australia's health, society, and economy. By bringing together the right kinds of data and making them easily accessible, researchers can answer a diverse range of research questions to inform policy and benefit public health. The potential impact of data-enabled health and medical research is enormous. One report estimates that the effective use of big data could deliver savings to national healthcare expenditure of around 8%, which would translate to more than $11 billion annually in Australia.[58] Some examples of opportunities for research areas that could be more easily advanced with improved data infrastructure include:

- Personalised and targeted healthcare
- Early detection of health vulnerabilities in individuals or communities
- Evaluating risk of hospitalisations and adverse outcomes
- Better identifying clinically relevant research or research of societal relevance
- Dynamic and nimble research in response to national and global trends
- Modelling and prediction of future health risks
- Advancing Aboriginal and Torres Strait Islander self-determined interests
- Understanding the impact of interventions in real-world populations and on long-term health outcomes
- Increasing research impact within the health system and performance tracking
- Reducing wasteful or unnecessary healthcare
- Making better use of digital technology for patient outcomes
- Linking knowledge across sectors for improved health outcomes
- Income from intellectual property
- Better understanding of preventive and therapeutic practices for health issues
- Better understanding gaps in care and public health interventions across population groups
- Empowering individuals to access their own health data to encourage appropriate preventative and therapeutic interventions

The scenarios developed for this project (Appendices C–E) asked contributors to imagine a future where health data research opportunities facilitated a landscape in which the consequences of dementia, obesity and AMR had reduced impact. Contributors were challenged to think outside the box and identify possible solutions for Australia's current data infrastructure gaps that could bring about this ambitious future.

While researchers were excited by the prospect of engaging in this kind of blue-sky thinking, it was clear from the discussions that the foundational gaps hindering simple, everyday tasks, were the most consequential, and could be addressed first, or at least simultaneously. As one researcher put it, "it is important not to try and put the cart before the horse" when developing solutions to benefit data-enabled health and medical research. The following section briefly proposes key short-, medium- and long-term suggestions that could be utilised to address the current gaps in Australia's health and medical research data landscape – combining these fundamental issues with more ambitious thinking and efforts to develop the workforce needed to underpin both.

Laying solid foundations by improving the basics of existing infrastructure

My Health Record

The Australian Government's My Health Record (MHR) is a valuable data resource for the health and medical research community. By the end of 2018, MHR was an opt-out system, including data sourced from both Commonwealth and state and territory collections for more than 90% of the population. Furthermore, MHR contains information that is very difficult to source for research purposes, such as pathology and radiology reports. Despite offering rich potential for the secondary use of health data, in reality MHR is currently not accessible for research. In addition, MHR data currently are not linked to outcome information (such as cause of death or coded hospital diagnoses) and include data in formats (such as free text) that cannot be easily utilised by researchers. "Research-ready" MHR datasets, with appropriate streamlined governance, would potentially have huge utility for research. If data principles such as FAIR and CARE were used to manage these data, the potential for advantages for health and medical research in the public interest to benefit patients could be significant.

Coordination and cohesion of existing data infrastructure, assets and policies

In the disparate Australian data landscape, improving the coordination and cohesion of current processes around governance, ethics and coordination between state and federal systems would provide researchers with the consistency they need to conduct high quality data-enabled research. The provision of standardised resources such as templates and example frameworks would also be invaluable to fast-track research. In addition, investment priorities should be directed towards improving data linkage between state and national data sets, as well as across services and sectors that impact health. The development of national policies that promote the harmonisation and streamlining of health data collection, linkage and sharing within existing data infrastructure could be prioritised for maximum impact to researchers.

Aboriginal and Torres Strait Islander data

Existing efforts to put research and data in the hands of Aboriginal and Torres Strait Islander peoples pursuing their self-determined interests should be expedited through partnership and collaboration, to ensure that the impact of current data gaps on the health equity of Aboriginal and Torres Strait Islander peoples are appropriately mitigated.

Building ambitious, long-term goals for future data infrastructure

"Plug and play" national data spine

An aspirational vision for a new model data asset that can address a number of the key gaps in existing infrastructure is that of a "plug and play" national data spine. This spine would be one large national data asset that links all major existing health data across states, territories and national assets, plus new data from registries and research studies. At present, researchers are encouraged to share their data to inform further research. However, the current infrastructure makes this challenging. The "plug and play" national data spine could provide the necessary mechanisms to achieve this goal. In addition, a key feature of this model would be that it could link these research data sets to government data sets, allowing researchers to capitalise on both of these important sources and maximise benefits from both directions. Data in this system would be dynamic, nimble, secure and private, and meet both the FAIR and CARE principles. The ‘spine’ would also be able to draw in cross-sector data relevant to health and could therefore be utilised by interdisciplinary researchers tackling challenges across Australia’s entire health landscape.

In the development of this new data asset, it would be essential to consider and address most of the data gaps, challenges and priorities noted in other sections of this report. It would not be possible to successfully implement a “plug and play” national data spine without first managing the current issues around data governance, ethics, privacy, linkage, sharing and more to create a cohesive, coordinated, and comprehensive foundation for such an asset. In this way, the spine represents an ambitious but holistic solution for the key issues facing health and medical researchers in Australia.

A key message from the AMR roundtable discussions indicated an appetite for a new “data lake” (a central location that holds a large amount of data, usually in its native or raw format). This “data lake” would securely hold large amounts of multidisciplinary and cross-sector data that could be easily accessed by researchers, government departments or industry on an anonymised basis. A “plug and play” national data spine, containing standardised, harmonised national data, could contribute “research ready” health data for integration with multisectoral data, helping tackle the issues noted in box 3.

All Australians as potential research subjects

In Australia, the vast majority of people with certain diseases, such as leukemia and breast cancer, who enter into treatment have the opportunity to become research subjects and often also see themselves as research subjects. Australia has the potential to broaden this culture of engaging with research to encompass all Australians. This is a significant challenge involving multiple complex factors over a substantial period of time. However, if this could be achieved, the collection of rich, robust and complete data could be utilised by research across all health and medical disciplines for the collective benefit of the Australian community. Data infrastructure, assets, policies and skills that facilitate this vision could have enormous benefits for data-enabled health and medical research. For example, infrastructure that facilitates opt-out rather than opt-in participation frameworks could be useful.

Harnessing electronic medical records

Ever larger volumes of more highly detailed clinical data are routinely captured in electronic form in the comprehensive electronic medical record (EMR) systems that are now operating in our public and private hospital systems, as well as in GP and other ambulatory care settings. Various initiatives around the country are creating hospital EMR “data lakes” and aggregated collections of primary care EMRs and MHR bring together extracts from multiple EMR sources for individuals. However, research using these data is in its infancy. Major barriers include lack of semantic interoperability– that is, a consistent way of encoding and representing clinical and other concepts – and the need to transform and encode data such as clinical text and images into analysable form. Investment in research efforts to address these challenges at national scale could accelerate EMR-based research and resultant benefits for patient care.

Artificial Intelligence (AI)

AI has the potential to help address important health challenges, as it can be used to analyse and identify patterns in large and complex data sets faster and more precisely than has previously been possible. AI built using multimodal EMRs and deep learning models is a dynamic research frontier and is already demonstrating vast potential to augment and improve diagnosis and personalise and optimise health care. AI has been shown to improve on standard clinical-based assessments, particularly in the field of image analysis, and can also be used to search the scientific literature for relevant studies, link information and combine different kinds of data. There are a number of important ethical and governance issues around the use of AI systems (many of which are highlighted in previous reports by AAHMS and ACOLA), and AI is also limited by the quality of the available health data.[59,60] However, in the future it is likely that AI systems will become more advanced and attain the ability to carry out a wider range of tasks with less human input. Investing in these systems now could substantially contribute to enhancing data-enabled research in the future. In addition, governments around the world are already investing heavily in their AI systems for data and while Australia is currently ranked 10th in the world for research quality in AI, we risk being left behind without continued advancement of this agenda.[61]

Building a workforce

The development and utilisation of a data-enabled health and medical research landscape will require a skilled, engaged and capable workforce. The gaps in Australia's data skills have been mentioned throughout this report and bear repeating given the depth and urgency of the problem. Investing in training programs that grow and retain a workforce with the necessary data skills and ensuring Australia can attract and retain international talent to supplement our expertise. This may include the prioritisation of appropriate visa provision to practitioners of these skills, as well as the creation of attractive career pathways within health and medical research.

8. The role of the Australian Academy of Health and Medical Sciences

As an impartial, authoritative, cross-sector voice in health and medical research, the Australian Academy of Health and Medical Sciences is well placed to support the advancement of data-enabled research in Australia.

The Academy can take specific steps to:

- Utilise the expertise and experience of our Fellows to highlight the importance and benefits of data-intensive research and the associated opportunities to improve health, and to support and advise the Australian community, including governments, on the development of Australia's national data infrastructure, policies and skills.
- Collaborate with Australia's other Learned Academies, ACOLA, ARDC and other key organisations to develop and inform national data infrastructure, policies and skills.
- Through our Fellowship, promote and encourage the adoption and implementation of the FAIR and CARE principles in data management.
- Periodically review Australia's progress towards an improved health and medical data landscape through consultation with Academy Fellows and the broader health and medical research community.
- Through our Fellowship, promote the ethical conduct of research including the appropriate use of data to meet research integrity and ethics standards in Australia and internationally.
- Through our Fellowship, promote transparent dialogue about the benefits of the use of health and medical data for research, and the attendant risks and ways to mitigate these.
- Through our Fellowship, promote and support the growth and consolidation of workforce expertise and academic disciplines that underpin health data science and those that are needed to ensure that effective use of data.

Appendix A: Health and Medical Data Steering Committee

The Academy is most grateful for the expert advice and guidance of the Health and Medical Data Steering Committee, which oversaw this project.

Chair

Professor Louisa Jorm FAHMS, Director, Centre for Big Data Research in Health, University of New South Wales

Committee members

Professor Melanie Bahlo FAHMS, Laboratory Head and Leader, Healthy Development and Ageing Theme, WEHI

Professor Julie Byles AO FAHMS, Director, Centre for Women's Health Research; Global Innovation Chair in Responsive Transitions in Health and Ageing, The University of Newcastle

Professor John Carlin FAHMS, Professor & Head, Clinical Epidemiology and Biostatistics Unit (CEBU), Murdoch Children's Research Institute; Professorial Fellow, Centre for Epidemiology & Biostatistics, The University of Melbourne

Professor Annette Dobson AM FAHMS, Professor of Biostatistics, School of Public Health, Faculty of Medicine, The University of Queensland

Professor Stephen Jan FAHMS, Co-Director, Health Systems Science; Head, Health economics and Process Evaluation Program, The George Institute for Global Health; Professor, Health Economics, Faculty of Medicine, UNSW; Honorary Professor, Sydney Medical School, University of Sydney

Professor Anton Peleg FAHMS, Professor, Infectious Diseases and Microbiology; NHMRC Practitioner Fellow; Director, Department of Infectious Diseases; Program Director, Alfred Infectious Diseases Program, the Alfred Hospital and Monash University

Secretariat

Lanika Mylvaganam, Policy Officer (lead secretariat)

Dr Amanda Rush, Policy Manager

The Academy is also grateful for the support for this project form Dr Cath Latham, Interim AAHMS CEO (until September 2021) and Catherine Luckin, AAHMS CEO (from October 2021).

Appendix B: Examples of Australia’s current health data infrastructure

The examples below represent only a handful of the available data infrastructure for health and medical research. They have been selected as representing a range of data examples, not specific to any disease, that are used by researchers in the health and medical sciences.

Resource/Platform	Description
Medical Benefits Schedule (MBS)	The MBS data collection contains information on services that qualify for a benefit under the Health Insurance Act 1973 and for which a claim has been processed. The database comprises information about MBS claims (including benefits paid), patients and service providers.[28]
Pharmaceutical Benefits Scheme (PBS)	The PBS data collection contains information on prescription medicines that qualify for a benefit under the National Health Act 1953 and for which a claim has been processed. The database comprises information about PBS scripts and payments, patients, prescribers, and dispensing pharmacies.[62]
National Mortality Database	This database holds records for deaths in Australia from 1964. It comprises information about multiple causes of death and other characteristics of the person such as age at death, sex, area of usual residence, ethnicity. [63]
National Health Survey	This survey collects data on Australia’s health characteristics including prevalence of long-term health conditions, health risk factors, demographic, and socioeconomic characteristics.[64]
The 45 and Up Study	A large-scale prospective cohort study of over 260,000 men and women aged 45 and over in NSW at recruitment in 2006-2009. Questionnaire data are linked to routinely collected datasets including death, hospital, cancer registration, MBS, and PBS datasets.[65]
State and territory government administrative data	Governments collect a range of data from different sources such as hospital data, registry data, surveys, and programs.
Population Health Research Network (PHRN)	PHRN supports and coordinates health data linkage infrastructure in Australia.[66]
The WA Data Linkage System (WADLS) The Centre for Health Record Linkage (CHeReL) The Centre for Victorian Data Linkage (CVDL) The Tasmanian Data Linkage Unit (TDLU) SA NT DataLink Data Linkage Queensland (DLQ) Australian Institute of Health and Welfare (AIHW) Data Integration Services Centre	These centres provide data linkage services and infrastructure to Australian researchers. They are also part of the PHRN.[29]
The National Integrated Health Services Information (NIHSI) Analysis Asset	This is a national asset which contains de-identified data from 2010-11 onwards on admitted patient care services (in public and private hospitals where available), emergency department services and

	outpatient services in public hospitals for all participating states and territories, along with MBS data, PBS and Repatriation Pharmaceutical Benefits Scheme data, Residential Aged Care data and National Deaths Index (NDI) data.[67]
Multi-Agency Data Integration Project (MADIP)	MADIP is a secure data asset combining information on health, education, government payments, income and taxation, employment, and population demographics over time.[68]
The Australian Research Data Commons (ARDC)	ARDC enables the Australian research community and industry access to national significant, data-intensive digital research infrastructure, platforms, skills, and collections of data.[69]
Australian BioCommons	This Commons aims to deliver digital support for Australian research on the molecular basis of life across environmental, agriculture and biomedical sciences.[70]
Currently under development	
My Health Record	My Health Record (MHR) is now an opt out service. Included information can range from pathology reports, diagnostic imaging reports, health history, medications, specialist letters and more.[71] MHR is not currently accessible for research despite the release of the ‘Framework to guide the secondary use of My Health Record system data’ in 2018. As yet, the MHR Secondary Use of Data Governance Board has not been established.
National Primary Health Care Data Asset	This asset is currently under development and will include information to better understand a patient’s journey and experiences within the primary healthcare system.[72]
The Health Studies Australian National Data Asset (HeSANDA)	This will be a national asset from the data outputs of health research projects.[73]
The LINked Data Asset for Australian Health Research (LINDAHR)	This asset will be designed to enable the progressive linkage and integration of high-value investigator-initiated health datasets, including registries, cohort studies and clinical trials from source datasets. The data asset framework, governance and infrastructure will enable the inclusion of longitudinal person-based records for hospitalisations, emergency department and aged care services, Medicare services, subsidised medicines, disability, notifiable conditions, cancers, cancer screening, deaths, and social determinants of health.[74]

Table 1: examples of Australia’s current health data infrastructure

References

1. Australian Institute of Health and Welfare. *Australia's Health 2020: Data Insights.*; 2020. Accessed January 24, 2022. https://www.aihw.gov.au/getmedia/be95235d-fd4d-4824-9ade-34b7491dd66f/aihw-aus-231.pdf.aspx?inline=true
2. Suter-Crazzolara C. Better patient outcomes through mining of biomedical big data. *Frontiers in ICT*. 2018;5(DEC):30. doi:10.3389/FICT.2018.00030/BIBTEX
3. EMC Corporation. *Digital Universe Driving Data Growth in Healthcare.*; 2014. Accessed January 25, 2022. https://www.cycloneinteractive.com/cyclone/assets/File/digital-universe-healthcare-vertical-report-ar.pdf
4. Auffray C, Balling R, Barroso I, et al. Making sense of big data in health research: Towards an EU action plan. *Genome Medicine*. 2016;8(1):1-13. doi:10.1186/S13073-016-0323-Y/FIGURES/1
5. Australian Institute of Health and Welfare. *Australia's Health 2018.*; 2018. Accessed January 24, 2022. https://www.aihw.gov.au/getmedia/57ed4b65-5919-43ce-bb21-933ea9a8b012/aihw-aus-221-chapter-2-5.pdf.aspx
6. Durcevic Sandra. 18 Examples of Big Data In Healthcare That Can Save People. Published 2020. Accessed January 25, 2022. https://www.datapine.com/blog/big-data-examples-in-healthcare/
7. Pastorino R, de Vito C, Migliara G, et al. Benefits and challenges of Big Data in healthcare: an overview of the European initiatives. *The European Journal of Public Health*. 2019;29(Suppl 3):23. doi:10.1093/EURPUB/CKZ168
8. Hernandez I, Zhang Y. Using predictive analytics and big data to optimize pharmaceutical outcomes. *American Journal of Health-System Pharmacy*. 2017;74(18):1494-1501. doi:10.2146/AJHP161011
9. Brook EL, Rosman DL, Holman CDAJ. Public good through data linkage: Measuring research outputs from the Western Australian Data Linkage System. *Australian and New Zealand Journal of Public Health*. 2008;32(1):19-23. doi:10.1111/J.1753-6405.2008.00160.X
10. Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. *Health Information Science and Systems*. 2014;2(1). doi:10.1186/2047-2501-2-3
11. Wilkinson MD, Dumontier M, Aalbersberg IjJ, et al. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data 2016 3:1*. 2016;3(1):1-9. doi:10.1038/sdata.2016.18
12. Wang X, Williams C, Liu ZH, Croghan J. Big data management challenges in health research—a literature review. *Briefings in Bioinformatics*. 2019;20(1):156-167. doi:10.1093/BIB/BBX086

13. Palamuthusingam D, Johnson DW, Hawley C, Pascoe E, Fahim M. Health data linkage research in Australia remains challenging. *Internal Medicine Journal*. 2019;49(4):539-544. doi:10.1111/IMJ.14244

14. Rainie SC, Kukutai T, Walter M, Figeuroa-Rodriguez O, Walker J, Axelsson P. Issues in Open Data - Indigenous Data Sovereignty. In: Davies T, Walker S, Rubinstein M, Perini F, eds. *The State of Open Data: Histories and Horizons*. African Minds and International Development Research Centre; 2019. Accessed January 25, 2022. https://www.stateofopendata.od4d.net/chapters/issues/indigenous-data.html

15. Walter M. The voice of Indigenous data - Beyond the markers of disadvantage. Accessed January 25, 2022. https://www.griffithreview.com/articles/voice-indigenous-data-beyond-disadvantage/

16. Dementia Australia. What is dementia? Accessed February 14, 2022. https://www.dementia.org.au/about-dementia/what-is-dementia

17. Royal Commission into Aged Care Quality and Safety. *Aged Care Reform: Projecting Future Impacts.*; 2020. Accessed January 27, 2022. www.dpmc.gov.au/government/commonwealth-coat-arms

18. Brown L, Hansnata E, La HA. *ECONOMIC COST OF DEMENTIA IN AUSTRALIA 2016 - 2056.*; 2017. Accessed January 27, 2022. http://www.natsem.canberra.edu.au/

19. Australian Institute of Health and Welfare. Overweight and obesity . Accessed February 14, 2022. https://www.aihw.gov.au/reports/australias-health/overweight-and-obesity

20. Australian Institute of Health and Welfare. Overweight and obesity: an interactive insight. Published 2020. Accessed January 27, 2022. https://www.aihw.gov.au/reports/overweight-obesity/overweight-and-obesity-an-interactive-insight/contents/what-is-overweight-and-obesity

21. Australian Institute of Health and Welfare. Australia's children. Published 2020. Accessed January 27, 2022. https://www.aihw.gov.au/reports/children-youth/australias-children/contents/health/overweight-and-obesity

22. Obesity Evidence Hub. Impacts. Accessed January 27, 2022. https://www.obesityevidencehub.org.au/collections/impacts

23. Okunogbe A, Nugent R, Spencer G, Ralston J, Wilding J. Economic impacts of overweight and obesity: current and future estimates for eight countries. *BMJ Glob Health*. 2021;6(10). doi:10.1136/BMJGH-2021-006351

24. Australian Commission on Safety and Quality in Health Care. *AURA 2019: Third Australian Report on Antimicrobial Use and Resistance in Human Health.*; 2019. Accessed January 27, 2022. https://www.safetyandquality.gov.au/sites/default/files/2019-06/AURA-2019-Report.pdf

25. Murray CJ, Ikuta KS, Sharara F, et al. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. *The Lancet*. 2022;399(10325):629-655.

doi:10.1016/S0140-6736(21)02724-0/ATTACHMENT/2CA9AF18-3728-4446-B9C7-CCE47BAFEB4F/MMC1.PDF

26. University of Technology Sydney. OUTBREAK: A One Health antimicrobial resistance economic perspective. Published 2020. Accessed January 27, 2022. https://outbreakproject.com.au/wp-content/uploads/2020/12/OUTBREAK_REPORT_2020_economics_ERRATUM.pdf

27. Food and Agriculture Organisation of the United Nations. Antimicrobial Resistance in the Environment. Published online 2018. Accessed January 27, 2022. https://www.fao.org/3/BU656en/bu656en.pdf

28. Department of Health. MBS online . Accessed February 1, 2022. http://www.mbsonline.gov.au/internet/mbsonline/publishing.nsf/Content/Home

29. Population Health Research Network. *Annual Review 2019-20*.; 2020. Accessed January 21, 2022. https://www.phrn.org.au/media/82004/phrn-annual-review-2019-2020_digital-v10.pdf

30. Department of Education S and E. *2021 National Research Infrastructure Roadmap* .; 2022. Accessed June 15, 2022. https://www.dese.gov.au/national-research-infrastructure/resources/2021-national-research-infrastructure-roadmap

31. Australian Academy of Science T. Advancing data-intensive research in Australia. Published online 2021. Accessed January 19, 2022. www.science.org.au

32. Lovett R, Walter M, Bodkin-Andrews G, Lee V. Maiam Nayri Wingara. Published 2017. Accessed February 1, 2022. https://www.maiamnayriwingara.org/

33. National Skills Commission. *The State of Australia's Skills 2021: Now and into the Future*.; 2021. Accessed February 2, 2022. https://www.nationalskillscommission.gov.au/sites/default/files/2021-12/2021%20State%20of%20Australia%27s%20Skills.pdf

34. Jorm L. Routinely collected data as a strategic resource for research: priorities for methods and workforce. *Public Health Res Pract*. 2015;25(4):e2541540. doi:10.17061/phrp2541540

35. Lee KJ, Moreno-Betancur M, Kasza J, Marschner IC, Barnett AG, Carlin JB. Biostatistics: a fundamental discipline at the core of modern health data science. *The Medical Journal of Australia*. 2019;211(10):444-446.e1. doi:10.5694/MJA2.50372

36. Queensland Genomics, Qeensland Government. *Blueprint for a National Approach to Genomic Information Management* .; 2020. Accessed March 1, 2022. https://queenslandgenomics.org/qldgenomics-updated/wp-content/uploads/2021/02/NAGIM-Blueprint-v20201010-Final-v1.2.2.pdf

37. Wilkinson MD, Dumontier M, Aalbersberg IjJ, et al. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data 2016 3:1*. 2016;3(1):1-9. doi:10.1038/sdata.2016.18

38. NHMRC. Open Access Policy . Published 2021. Accessed January 31, 2022. https://www.nhmrc.gov.au/about-us/resources/open-access-policy

39. ARDC. FAIR Data Guidelines for Project Data Outputs . Accessed January 31, 2022. https://ardc.edu.au/about_us/policies-and-guidelines/fair-data-guidelines-for-project-data-outputs/

40. Srinivasan U, Ramachandran D, Quilty C, Rao S, Nolan M, Jonas D. *Flying Blind: Australian Researchers and Digital Health*. Vol 2. Digital Health Cooperative Research Centre; 2018. Accessed January 31, 2022. https://apo.org.au/node/206591

41. Global Indigenous Data Alliance. CARE Principles of Indigenous Data Governance. Accessed January 31, 2022. https://www.gida-global.org/care

42. Carroll SR, Garba I, Figueroa-Rodríguez OL, et al. The CARE principles for indigenous data governance. *Data Science Journal*. 2020;19(1):1-12. doi:10.5334/DSJ-2020-043/METRICS/

43. Menzies Institute for Medical Research. What is Data Linkage? Accessed January 31, 2022. https://www.menzies.utas.edu.au/research/research-centres/data-linkage-unit/what-is-data-linkage

44. Young A, Flack F. Recent trends in the use of linked data in Australia. *Australian Health Review*. 2018;42:584-590. doi:10.1071/AH18014

45. Mitchell RJ, Cameron CM, McClure RJ, Williamson AM. Data linkage capabilities in Australia: Practical issues identified by a Population Health Research Network "Proof of Concept project." *Australian and New Zealand Journal of Public Health*. 2015;39(4):319-325. doi:10.1111/1753-6405.12310

46. Shaw JA, Sethi N, Cassel CK. Social license for the use of big data in the COVID-19 era. *npj Digital Medicine 2020 3:1*. 2020;3(1):1-3. doi:10.1038/s41746-020-00342-y

47. Muller SHA, Kalkman S, van Thiel GJMW, Mostert M, van Delden JJM. The social licence for data-intensive health research: towards co-creation, public value and trust. *BMC Medical Ethics*. 2021;22(1):1-9. doi:10.1186/S12910-021-00677-5/TABLES/1

48. Department of Education Skills and Employment. 2016 National Research Infrastructure Roadmap. Published 2017. Accessed February 1, 2022. https://www.dese.gov.au/national-research-infrastructure/resources/2016-national-research-infrastructure-roadmap

49. Paterson M, Witzleb N. The Privacy-Related Challenges Facing Medical Research in an Era of Big Data Analytics: A Critical Analysis of Australian Legal and Regulatory Frameworks . *J Law Med*. 2018;26(1):188-203. Accessed January 25, 2022. https://pubmed.ncbi.nlm.nih.gov/30302981/

50. Australian Institute of Health and Welfare. The Five Safes framework. Published 2021. Accessed March 15, 2022. https://www.aihw.gov.au/about-our-data/data-governance/the-five-safes-framework

51. Research Australia. *2020 Public Opinion Poll on Health and Medical Research & Innovation.*; 2020. Accessed January 21, 2022. https://issuu.com/researchaustralia/docs/2020_opinion_poll_report_final

52. de Lusignan S, van Weel C. The use of routinely collected computer data for research in primary care: opportunities and challenges. Published online 2005. doi:10.1093/fampra/cmi106

53. Jones B, Heslop D, Harrison R. Seldom heard voices: a meta-narrative systematic review of Aboriginal and Torres Strait Islander peoples healthcare experiences. *International Journal for Equity in Health*. 2020;19(1):1-11. doi:10.1186/S12939-020-01334-W/FIGURES/1

54. Javanparast S, Naqvi SKA, Mwanri L. Health service access and utilisation amongst culturally and linguistically diverse populations in regional South Australia: a qualitative study. *Rural Remote Health*. 2020;20(4). doi:10.22605/RRH5694

55. Knight HE, Deeny SR, Dreyer K, et al. Challenging racism in the use of health data. *The Lancet Digital Health*. 2021;3(3):e144-e146. doi:10.1016/S2589-7500(21)00019-4

56. Brodie MA, Pliner EM, Ho A, et al. Big data vs accurate data in health research: Large-scale physical activity monitoring, smartphones, wearable devices and risk of unconscious bias. *Medical Hypotheses*. 2018;119:32-36. doi:10.1016/J.MEHY.2018.07.015

57. Parliament of Australia. Data Availability and Transparency Bill 2020. Published 2020. Accessed February 2, 2022. https://www.aph.gov.au/Parliamentary_Business/Bills_LEGislation/Bills_Search_Results/Result?bId=r6649

58. Centre for Big Data Research in Health. *Annual Report 2019.*; 2019. Accessed February 2, 2022. https://cbdrh.med.unsw.edu.au/sites/default/files/UN1631%20UNSW%20GSR%20-%20CBDRH%20A_REPORT_digital_single_1.pdf

59. ACOLA. The effective and ethical development of artificial intelligence: An opportunity to improve our wellbeing . Published 2020. Accessed February 14, 2022. https://acola.org/hs4-artificial-intelligence-australia/

60. The Australian Academy of Health and Medical Sciences. RESPONSE TO THE DEPARTMENT OF INDUSTRY, INNOVATION AND SCIENCE CONSULTATION ON THE ARTIFICIAL INTELLIGENCE: AUSTRALIA’S ETHICS FRAMEWORK, MAY 2019. Published 2019. Accessed February 14, 2022. https://aahms.org/wp-content/uploads/2019/06/AAHMS_Consultation-Response_Artificial-Intelligence-Australias-Ethics-Framework.pdf

61. Australian Academy of Science, Australian Academy of Engineering and Technology. Preparing for Australia’ s Digital Future A strategic plan for information and communications science, engineering and technology . Published online 2019. Accessed February 3, 2022. www.science.org.au/digital-future

62. Australian Government. Pharmaceutical Benefits Scheme (PBS) - Item Report - Dataset. Accessed February 1, 2022. https://data.gov.au/data/dataset/pharmaceutical-benefits-scheme-pbs-item-report

63. Australian Institute of Health and Welfare. National Mortality Database (NMD). Accessed February 1, 2022. https://www.aihw.gov.au/about-our-data/our-data-collections/national-mortality-database

64. Australian Bureau of Statistics. National Health Survey: First results, 2017-18 financial year . Published 2018. Accessed February 1, 2022. https://www.abs.gov.au/statistics/health/health-conditions-and-risks/national-health-survey-first-results/latest-release

65. The Sax Institute. *Accessing the 45 and Up Study*.; 2020. Accessed February 15, 2022. www.saxinstitute.org.au

66. Population Health Research Network. PHRN. Accessed February 1, 2022. https://www.phrn.org.au/

67. Australian Institute of Health and Welfare. Approved AIHW linkage projects . Accessed February 2, 2022. https://www.aihw.gov.au/our-services/data-linkage/approved-aihw-linkage-projects

68. Australian Bureau of Statistics. Multi-Agency Data Integration Project (MADIP) . Accessed March 1, 2022. https://www.abs.gov.au/about/data-services/data-integration/integrated-data/multi-agency-data-integration-project-madip

69. Australian Research Data Commons. ARDC. Accessed February 1, 2022. https://ardc.edu.au/

70. Australian BioCommons. Australian BioCommons. Accessed February 1, 2022. https://www.biocommons.org.au/

71. Australian Digital Health Agency. My Health Record. Accessed February 1, 2022. https://www.myhealthrecord.gov.au/

72. Australian Institute of Health and Welfare. Primary health care data development. Accessed February 1, 2022. https://www.aihw.gov.au/reports-data/health-welfare-services/primary-health-care/primary-health-care-data-development

73. ARDC. Health Studies National Data Asset program. Accessed January 19, 2022. https://ardc.edu.au/collaborations/strategic-activities/national-data-assets/health-studies-national-data-asset-program/

74. Australian Research Data Commons. LINked Data Asset for Australian Health Research (LINDAHR). Published 2021. Accessed February 3, 2022. https://ardc.edu.au/project/linked-data-asset-for-australian-health-research-lindahr/

75. Commonwealth of Australia 2022. *The National Obesity Strategy 2022-2032*.; 2022. Accessed March 8, 2022.

https://www.health.gov.au/sites/default/files/documents/2022/03/national-obesity-strategy-2022-2032.pdf

76. Australian Institute of Health and Welfare. *Dementia Data Gaps and Opportunities*.; 2020. doi:10.25816/xmer-mr26