Cancer Council

Developing a National Data Strategy for Cancer: A Report for Discussion

October 2023

Cancer Council is the peak, non-Government cancer control organisation in Australia. As the national body in a federation of eight state and territory member organisations, Cancer Council Australia works to make a lasting impact on cancer outcomes by: shaping and influencing policy and practice across the cancer control continuum; developing and disseminating evidence-based cancer information; convening and collaborating with cross sectorial stakeholders and consumers to set priorities; and speaking as a trusted voice on cancer control in Australia.

Cancer Council acknowledge the traditional custodians of the lands on which we live and work. We pay respect to Aboriginal and Torres Strait Islander elders past, present and emerging and extend that respect to all other Aboriginal and Torres Strait Islander people.

ISBN 978-0-6457294-1-2

Published October 2023

Prepared and produced by Cancer Council Australia

Megan Varlow
Direct Cancer Control Policy, Cancer Council Australia
E: Megan.Varlow@cancer.org.au
T: 02 8063 4111

cancer.org.au

Developing a National Data Strategy for Cancer: A Report for Discussion

Contents

Project team: Dr Belinda Goodwin, Kate Whittaker, Professor Peter Baade, Professor Joanne Aitken, Megan Varlow, Professor Sanchia Aranda AM, Professor Sue Evans.

Overview

Cancer Council's vision is for Australia to have a robust performance measurement and reporting system that holds all stakeholders (governments, payers, health service executives, clinicians) to account for the delivery of optimal treatment and care for people affected by cancer. Such a system must be national in scope, allow comparisons adjusted for patient and disease characteristics and reported publicly to enable informed choice by people affected by cancer. It would drive efforts to improve outcomes overall, support closing the gaps experienced by disadvantaged populations and lead to a more efficient and cost-effective health system. A continual process for monitoring, evaluating and communicating knowledge about the performance of the health services and their progress towards targets could lead to achieving and maintaining the delivery of safe, high-quality cancer care and experiences.

While Australia has an exceptional health system and cancer outcomes for the population are among the best in the world,[1] cancer survival varies depending on factors such as ethnic group, socio-economic status, education level, residential location, and where and from whom individuals receive healthcare.[2,3,4] At a national level, determining the best intervention approaches to close gaps in cancer outcomes is hampered by poor availability and accessibility of data at a population level.[5,6] In addition, analyses of available treatment data in Australia demonstrate significant unexplained variation in clinical practice, with limited capacity to analyse the impact of this variation on patient outcomes.[7] Most of the data collected from and about people with cancer in Australia are neither available nor utilised to improve health care and system performance. Australia has no standardised or mandatory performance measurement and reporting system for health services and is therefore missing the information needed to benchmark performance across the system and to systematically identify opportunities for improvement. It is up to individual health services, networks, or jurisdictions to adopt reporting mechanisms and then to make this publicly available. In some cases, data have been collected for decades but never collated or reported back to health services and clinicians for the purpose of improving outcomes.[7] Additionally, patient reported measures, which are among the most important measures in assessing whether or not healthcare is delivering value to patients, are not captured at a national level.[8] Combined, this impedes our collective ability to maximise use of existing data and to create new insights that will assist in improving cancer outcomes and managing health-service costs.[6]

To define a vision for improving cancer data in Australia, international and national best practice approaches to data availability and access were reviewed, and, in consultation with key stakeholders, a data maturity model for Australia was developed. This describes a data maturity strategy for cancer data that can be applied in the Australian context, detailing the current situation, the ideal vision, and the steps to achieve this vision. A mature data system would have the ability to produce a routinely updated national master linked dataset fed by State, Territory and Commonwealth data collections, held in a secure research environment accessible by accredited and accountable analysts and researchers, with simplified data custodianship and further capacity to link collections such as quality clinical registries. Achieving a mature cancer data system will facilitate a National Health System Performance Reporting Framework which could support performance assessment, planning and benchmarking within the health sector. A mature data system could use the high-quality, scientific evidence that facilitates improved cancer outcomes to support real-time health system reporting and population health research. Real-time health system monitoring would then enable Australia to benchmark against national and international cancer outcomes and improve service delivery by providing automatic real-time system monitoring and feedback to guide resources and attention. Tracking of the implementation and outcomes of prevention, screening and early detection activities would be improved and monitoring geographical and other disparities in outcomes could advance population health outcomes.

Cancer Council believes addressing barriers to improving the utilisation of health data in Australia should be a national priority, and that a data maturity model identifies opportunities to engage various stakeholders in improving collection and utilisation of cancer data in Australia. To achieve high-quality, scientific evidence that facilitates improved cancer outcomes in Australia from real-time health system monitoring and population health research, Australia must invest in both improving public recognition of the value of data and the infrastructure and systems which enable their efficient collection and use.

Vision for Cancer Data in Australia

High-quality, scientific evidence that facilitates improved cancer outcomes in Australia from:

REAL-TIME HEALTH SYSTEM MONITORING

- National and international benchmarking of cancer outcomes
- Automatic real-time system monitoring and feedback
- Post marketing surveillance

POPULATION HEALTH RESEARCH

- Monitoring of geographical and other disparities in outcome
- Ongoing follow up of patient cohorts
- Tracking implementation and outcomes of prevention/early detection initiatives

VISION FOR CANCER DATA MATURITY

CAPACITY

DATA ITEMS AND COLLECTION

- National Health Data Repository
- National Health Data Dictionary
- Structured clinical & pathology reporting
- Nationally applied unique linkage key

RESOURCES

- Adequate level of dedicated funding
- Qualified and trained workforce
- Implementation of artificial intelligence
- High-speed, high capacity, secure data portal for data transfer

DATA QUALITY

- Regular transparent auditing of data protocols, procedures and quality
- Efficient and effective infrastructure and processes for data quality assurance
- Data portal for data transfer

DATA LINKAGE AND TRANSFER

- Real-time enduring links between jurisdictional cancer incidence data and all other relevant population-wide datasets including hospitalisation, treatment, mortality
- All jurisdictional linked cancer data combined in a single National Health Data Repository
- Deterministic linkage through a national unique linkage key
- Collect once, link once, and use multiple times
- Macros/syntax shared and/or aggregated and cleaned data feedback into the dataset

ANALYTIC CAPABILITY

- Complex statistical analytic capability
- Big data analytic capability
- Real time automated analyses and reporting

ACCESS

CULTURE

- Strong community, stakeholder and political trust in and support for routine health data collection for ongoing monitoring and quality improvement of the health system (no opt-out option)
- Transparent and public reporting on what data held, how data security is maintained and how data are being used to improve health outcomes

GOVERNANCE

- Nationally uniform legislation and guidelines around rights and responsibilities of data custodians and users that cover the collection, linkage, use, access and dissemination data for research
- Legislation that allows the use of data collected for the benefit of patient care to be used to improve patient care without consent or ethics approval

CUSTODIANSHIP

- Single national independent data custodian for the National Health Data Repository
- Single, streamlined ethical and data custodian approval for use of nationally linked unit-level data for research
- Ethics approval not required for monitoring of health system performance

SECURITY

- Proportionate risk avoidance measures in place
- Agreed national protocols for safe, ethical, and effective collection and storage of data (e.g., five-safes principles enforced)

A Data Maturity Model for Cancer Data in Australia

The Data Maturity Model for Cancer Data in Australia describes the status of cancer data maturity in Australia and identifies the key steps for achieving the proposed vision of a National Health Data Repository and a corresponding National Health Data Dictionary. The National Health Data Repository is seen, not as a collection of mandated data items, but rather as a repository of all health data that are available on a population basis.

Method

Based on a scoping review of the literature, review of the existing data situation in Australia, review of international data systems (namely the Canadian and UK data systems), and preliminary investigations of the expert Data Working Group of Cancer Council Australia, a draft vision for cancer data maturity in Australia was developed. The draft model described the current status and specified an 'ideal state' for each of the proposed key elements of cancer data maturity and was distributed to key stakeholders for comment. Key stakeholders included data custodians, health departments, researchers including those with experience of Indigenous health data, and data linkage experts nationally and internationally.

Interviews with key stakeholders focussed on three key questions:

1. Do the components adequately describe cancer data maturity for Australia, and if not, what components would you change or add?
2. For each component, do you agree with the stated vision and if not, how would you change it?
3. What do you think are the barriers, challenges, or facilitators in moving from the current situation to data maturity?

Stakeholder feedback was considered in finalising the Data Maturity Model for Cancer Data in Australia.

Twelve data collections were identified as providing a broadly representative snapshot of the status of cancer-related data collections in Australia. These were Australian Cancer Database (ACD), BreastScreen Australia dataset, Estimated Resident Population dataset (Australia), Medicare Benefits Schedule (MBS), National Death Index (NDI), National Mortality Database (NMD), New South Wales Cancer Registry, New South Wales Hospital Admitted Patient Data Collection, Queensland Cancer Register, Queensland Hospital Admitted Patient Data Collection, Victorian Admitted Episodes dataset, and Victorian Cancer Registry.

Audits of the publicly available information on the website of these data collections were undertaken to determine the status of 28 individual characteristics relating to the nine categories within the Data Maturity Model for Cancer Data in Australia. The template for information captured for each dataset is included in the Supplementary Materials. The current status of each of the components of the proposed Cancer Data Maturity Model were captured, and the remaining steps required to realise the stated vision for data maturity for Australia were identified.


Cancer Council
Family, Friends & Carers
Contact
Sign in
Search
Menu
ly, Friends and Carers
SAMSUNG

A Data Maturity Model for Cancer Data in Australia

		CURRENT STATUS	STEP 1	STEP 2	DATA MATURITY
CAPACITY	**Data Items & Collection**	• Limited dataset definitions for some cancers • Some national standard definitions, but not universally applied • Establish unique person identifiers for national health data repository • METeOR, national data dictionary	• Obtain broad agreement on the need for national health data dictionary • Work with clinicians to improve and standardise clinical reporting consistent with national health data dictionary	• Build a comprehensive national health data dictionary • Mandate adherence to the national data dictionary by all relevant data collections	• National Health Data Repository • National Health Data Dictionary • Structured clinical and pathology reporting • Nationally applied unique linkage key
	Resources	• Technology and expertise exist • High speed, remote access, secure data portals currently in use	• Evaluate information technology needs • Evaluate gaps in personnel and relevant training • Secure adequate funding	• Resource and train personnel • Test and implement advanced information technology systems	• Adequate level of dedicated funding • Qualified and trained workforce • Implementation of artificial intelligence • High-speed, high capacity, secure data portal for data transfer
	Data Quality	• Regular data quality monitoring • Not consistent between data collections • No standard, universal protocols for reporting and resolving errors	• Establish and define standard protocols for data quality monitoring and error reporting	• Mandate adherence to defined quality monitoring and error reporting protocols	• Regular, transparent auditing of data protocols, procedures and quality • Efficient and effective infrastructure and processes for data quality assurance
	Data Linkage & Transfer	• Some linkage possible for individual projects • Some routine data linkage for government data collections • Timeliness of linked data is variable	• Establish unique person identifiers (linkage keys)	• Mandate the use of unique identifiers to all relevant data collections • Establish real-time jurisdictional cancer incidence data • Centralised cancer registry platform available to all jurisdictions which incorporates AI and NLP to extract data elements, data derivation and validation.	• Real-time enduring linkage between jurisdictional cancer incidence data and all other relevant population-wide datasets including hospitalisation, treatment, mortality • All jurisdictional linked cancer data combined in a single National Health Data Repository • Deterministic linkage through a national, unique linkage key • Collect once, link once, and use multiple times • Macros/syntax shared and/or aggregated and cleaned data fed back into the dataset
	Analytic Capability	• Routine analysis for health system monitoring at state level • Some routine descriptive analysis at national level • Basic tools available within researcher-accessible national linkages	• Identify the availability of individual analytic tools into secure environment on request	• Integrated individual analytic tools in the national health data repository	• Complex statistical analytic capability • Big data analytic capability • Real time automated analyses and reporting
ACCESS	**Culture**	• No broad community engagement or understanding of benefits • Limited transparency and consultation between stakeholders and custodians • Behaviour driven by risk aversion	• High public awareness of benefits • Data custodians and community leaders engaged in planning and discussions	• Formal support for the Data Maturity Model from State, Territory and Commonwealth Health Ministers	• Strong community, stakeholder and political trust in and support for routine health data collection for ongoing monitoring and quality improvement of the health system (no opt-out option) • Transparent and public reporting on what data are held, how data security is maintained and how data are being used to improve health outcomes
	Governance	• Multiple and inconsistent legislation between and within jurisdictions • Plain language documentation for users is provided but not always clear • Data custodians bound by restrictive legislation	• Specific legislative barriers identified • Recommendations for legislative change developed	• Formal State, Territory and Commonwealth support for required legislative changes	• Nationally uniform legislation and guidelines around rights and responsibilities of data custodians and users that cover the collection, linkage, use, access to and dissemination of data for research • Legislation that allows the use of data collected for the benefit of patient care to be used to improve patient care without consent or ethics approval
	Custodianship	• Separate data custodian for each data collection • Data custodians have legal responsibility and full control over siloed datasets • Access to linked data requires approval from each contributing data custodian	• Appoint a single jurisdiction-based data custodian for each state and territory • Design and document process for quicker ethics and access approval for linked datasets	• Appoint a single national independent data custodian for the National Health Data Repository	• Single national independent data custodian for the National Health Data Repository • Single, streamlined ethical and data custodian approval for use of nationally linked unit-level data for research • Ethics approval not required for monitoring of health system performance
	Security	• Reasonable harmonisation (most security protocols based on the Five Safes framework)	• Document a single national protocol for secure collection, storage and transfer of relevant data collections based on • Five-Safes framework[16]	• Mandate adherence to national protocol for secure collection, storage and transfer of data based on Five-Safes framework[16]	• Proportionate risk avoidance measures in place • Agreed national protocols for safe, ethical and effective collection and storage of data (e.g., Five-Safes principles)

Guiding principles for improving cancer data in Australia

Several principles emerged from the development of the Data Maturity Model and were workshopped by several experts. This preliminary set of guiding principles sit alongside the data maturity model and will evolve as Australia's data system becomes more mature.

Improvement of health and the health system is the primary purpose for data collection and use. Data are treated as a public good, not a by-product.

Data maturity is marked by a readiness to use data strategically and efficiently, not only in terms of resources, but also attitudes and culture. A mature data system is embedded in a culture where data resources are valued, shared and utilised with a primary goal of improving health and health system performance. The primary purposes of using health data are therefore multiple and varied.

A mature data system is driven by the purpose for which the data are used, improving health and the health system.

In a mature data system data are used routinely to inform decisions and to improve outcomes. Data systems should be designed to proactively integrate end user needs, to ensure data can be assessed and answer the necessary questions, and feed into national health priorities.

The purpose and use of data are transparent to all.

A mature data system requires the acceptance and collaboration of all stakeholders in the use of data to improve health and health system outcomes. Open reporting to the community on what data are held, how data security is maintained and how data are being used will foster acceptance, accountability, and ultimately facilitate efforts to improve health system performance and outcomes.

A mature data system should enable forecasting future demand, not simply describe current patterns.

A mature data system includes national, where feasible, standard and uniform, data definitions that address quality and are accessible to all. A mature data system is built on routine linkages between demographic, administrative, clinical, registry and research datasets.

Data are shared across the data system and used multiple times.

A mature data system that is fit for purpose makes efficient use of its data. A mature data system can share data into different datasets, removing unnecessary duplication of collection. Mechanisms must be adopted which incorporate additional data fields derived from the data back into the data system for use by all, and ensure reliable measures are maintained. Application of a "collect once, use multiple times" principle ensures timely and efficient utilisation of data.

Patient data are routinely used to improve their health and the healthcare they receive.

A mature data system uses quality point-of-care data to improve outcomes. A mature data system requires broad acceptance by the community that the collection and use of health data for an individual patient's care and system improvement is an inherent part of using the health care system. As such, people have a right to be cared for within the system, but a duty to contribute to the improvement. The more the data are entered in structured format, the more readily accessible it will be.

Resources are appropriate and dedicated to delivering and maintaining a mature data system long term.

To establish and maintain a mature data system, adequate resources for technology and personnel are vital. A commitment to build and foster a skilled and capable workforce is an important step in establishing and sustaining data maturity in Australia.

Data are shared securely, with privacy proactively embedded into system design.

Data transfer, sharing, and linkage are key in achieving the vision for data maturity. Data security and maintaining privacy are therefore of utmost importance. Data maturity requires a system in which privacy is embedded into the collection, storage and use of data, ensuring that personal information is both safe and accessible.

Identification and analysis of available datasets in Australia for cancer-related indicators

This work aimed to identify and describe population-based health- and cancer-related datasets in Australia that could be used to quantify disparities in cancer-related indicators. Eligible data collections included those population-based health- and cancer-related datasets that are currently held by jurisdictions, government agencies and research institutes within Australia and contain data that could potentially be used to identify and measure disparities in cancer indicators, either alone or through linkage to other datasets. While there is increasing appreciation of the value in using both health and non-health datasets (like those held by the Australian Bureau of Statistics) to understand the quality of cancer care, particularly in priority health populations, these datasets were not considered within this analysis.

Relevant health- and cancer-related datasets were identified, including their attributes such as data items, scope, and access requirements; and an assessment made of the capacity of these datasets both to identify and measure inequities in cancer outcomes and to inform the development of indicators and improvement measures.

Datasets were identified through web searches and guidance from expert informants (further detail can be provided on request). Datasets were included if they contained data that could potentially be used to identify and measure disparities in cancer outcomes, whether alone or through linkage to other datasets. Using a template, online descriptions of each dataset were reviewed, and the characteristics and relevant information of each dataset were recorded. The template is included in the Supplementary Materials. The custodians of each dataset were then contacted and asked to confirm the accuracy of the extracted metadata and to provided missing metadata that was not available on their website. A total of 25 datasets were identified and categorised according to their use. In instances where state and territories hold the data for their jurisdiction and these jurisdictional datasets are combined in a central national dataset, only the combined national datasets is described.

Datasets identified for review

Category	Datasets
Population cancer registry data	Australian Cancer Database.
Population administrative data	National Mortality Database; Mortality Enhanced Characteristics; Estimated Resident Population Dataset (Australia); Multi-Agency Data Integration Project (MADIP)*.
Health service billing data	Medicare Benefits Schedule Data Collection; Pharmaceutical Benefits Scheme (PBS) Data Collection.
Hospital and health service data	National Hospitals Dataset; Hospital Case mix Protocol; Private Hospital Data Bureau; Radiotherapy Waiting Times National Minimum Dataset; Palliative Care Outcomes Collaboration Data.
Screening data	BreastScreen Australia; National Cancer Screening Register (NCSR): National Bowel Cancer Screening Program.
Australian Bureau of Statistics	Australian Aboriginal and Torres Strait Islander Health Survey; Australian Census Longitudinal Data; National Nutrition and Physical Activity Survey; National Health Survey; Patient Experiences in Australia.
Research data	45 and Up Study; QSkin Study; Melbourne Collaborative Cohort Study (Health 2020); Australian Childhood Cancer Registry; Mesothelioma Registry; Prostate Cancer Outcomes Registry

*The metadata for MADIP were not able to be verified with data custodians prior to submitting this report and are therefore not included in the following summary table.

Datasets were examined for whether they included individual unit records, were population-based, include or can be linked to individual and area-level data, and include one or more cancer-related indicators of key interest including those relating to cancer diagnosis, treatment, or cancer mortality. The capacity of each dataset for record linkage was also assessed. A summary of the dataset review is provided overpage.

	1.	2.	3. Area and individual-level characteristics					4. Indicators (available or calculatable)											5.
	Unit- level records	**Population coverage**	Smallest geographical area	Sex	Indigenous status	Country of birth	Others	Cancer Diagnosis (Type)	Treatment	Stage	*Incidence*	*Mortality*	*Relative survival*	*Cause specific survival*	*Post Treatment Mortality*	Comorbidities	Other individual risk factors	Self-report health data	**Individual data linkage**
POPULATION CANCER REGISTRY DATA																			
Australian Cancer Database	✓	✓	S2	✓	-	✓	✗	✓	✗	✗	✓	✓	✓	✗	✗	✗	✗	✗	✓
POPULATION ADMINISTRATIVE DATA																			
National Mortality Database	✓	✓	S2	✓	✓	✓	✗	✗	✗	✗	✗	✓	✓	✓	✗	✗	✗	✗	✓
Mortality Enhanced Characteristics	✓	✓	S2	✓	✓	✓	✓	✗	✗	✗	✓	✓	✓	✓	✗	✗	✗	✗	✗
Estimated Resident Population Dataset	✗	✓	S2	✓	✓	✗	✓	✗	✗	✗	✗	✓	✗	✗	✗	✗	✗	✗	✗
HEALTH SERVICE BILLING DATA																			
Medicare Benefits Schedule (MBS) Data Collection	✓	✓	P	✓	✗	✗	✗	✗	-	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓
Pharmaceutical Benefits Scheme (PBS) Data Collection	✓	✓	P	✓	✗	✗	✗	✗	-	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓
HOSPITAL AND HEALTH SERVICE DATA																			
National Hospitals Dataset	-	✓	S2	✓	✓	✓	✗	-	✓	-	-	-	✗	✗	✗	✓	✗	✗	✓
Hospital Casemix Protocol	✓	✓	P	✓	✗	✗	✗	✓	✓	✗	✗	✗	✗	✗	✗	✓	✗	✗	✗
Private Hospital Data Bureau	✓	✓	P	✓	✗	✗	✗	✓	✓	?	✗	✗	✗	✗	✗	✓	✗	✗	✗
Radiotherapy waiting times	✓	✓	S2	✓	✓	✗	✗	✓	✗	✗	✓	✓	✗	✗	✗	✗	✗	✗	✗
Palliative Care Outcomes Collaboration	✓	✗	P	✓	✓	✓	✓	✓	✗	✗	✗	✗	✗	✗	✗	?	✗	✓	✓
SCREENING DATA																			
BreastScreen Australia	✓	✓	P	-	✓	✗	✓	S	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗	
National Cancer Screening Register	✓	✓	A	-	✓	✓	✓	S	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓
National Bowel Cancer Screening Program	✓	✓	A	✓	✓	✓	✓	S	✗	✗	✓	✗	✗	✗	✗	✗	✗	✗	✓
ABS SURVEY DATA																			
National ATSI Health Survey	✓	✗	R	✓	✓	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✓	✗
Australian Census Longitudinal Dataset	✓	✗	S2	✓	✓	✓	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
National Nutrition and Physical Activity	✓	✗	R	✓	✗	✓	✓	✗	✗	✗	✗	✗	✗	✗	✗	✓	✓	✓	✗
National Health Survey	✓	✗	R	✓	✓	✓	✓	✓	✓	✗	✗	✗	✗	✗	✗	✓	✓	✓	✗
Patient Experience Survey	✓	✗	S1	✓	✓	✓	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗
RESEARCH DATA																			
45 and Up Study	✓	✗	S2	✓	✓	✓	✓	✓	✓	✗	✓	✓	✗	✗	✗	✓	✓	✓	✓
QSkin Sun & Health Study	✓	✗	P	✓	✓	✓	✓	S	✗	✗	✓	✓	✓	✓	✗	✗	✓	✓	✓
Melbourne Collaborative Cohort Study	✓	✗	A	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	✓	✓	✓	✓	✓	✓
Australian Childhood Cancer Registry	✓	✓	S2	✓	✓	✓	✗	✓	✓	✓	✓	✓	✓	✓	✓	✗	✗	✗	✓
Australian Mesothelioma Registry	✓	✓	S2	✓	✓	✓	✓	S	✗	✗	✓	✓	✓	✓	✗	✗	✓	✓	✓
Prostate Cancer Outcome Registry	✗	✓	A	-	✗	✓	✗	S	✓	✓	✓	✓	✓	✓	✓	✗	✗	✗	✓

A = Address, **S2** = Statistical Area Level 2, **S1**= Statistical Area Level 1 **P** = Postcode, **HP** = health care providers, **s** = Specific Cancer, **R** = Remoteness (according to ASGS), **ATSI** = Aboriginal and Torres Strait Islander,
ABS = Australian Bureau of Statistics, ˜ = in some cases

Suggested Next Steps

A collaborative approach and national commitment is required to achieving a mature cancer data ecosystem supporting the strategic, efficient and effective management of data at all levels of the health system. This report is intended to provide a foundation for further discussion. We propose the following steps to progress this work:

- Identify stakeholders with whom to engage in progressing this strategy
- Engage with stakeholders to contribute to the proposed strategy so we can better understand:
  - the current landscape, including whether the datasets identified in Table 2 accurately capture existing datasets.
  - risks, including an analysis of the political, economic, sociological, technological, legal, environmental, ethical and demographic factors that are likely to impact success of the strategy.
  - opportunities that will emerge and value which will be realised as adoption moves across the various steps in the framework
- Prioritise and develop a comprehensive national data implementation strategy, which is costed and provides a timeframe for delivery across the steps.

This Report is intended to provoke discussion and provide a catalyst for change. It is ambitious and it is too early to definitively determine Australia's ability to achieve full data maturity. This will become evident as these next steps are undertaken.

References

1. Arnold M, Rutherford MJ, Bardot A, Ferlay J, Andersson TML, Myklebust TÅ, et al. Progress in cancer survival, mortality, and incidence in seven high-income countries 1995–2014 (icbp survmark-2): A population-based study. The Lancet Oncology. 2019;20(11):1493-505.
2. Bygrave A, Whittaker K, Aranda S. Inequalities in cancer outcomes by indigenous status and socioeconomic quintile: An integrative review. Sydney (AU); 2020. https://www.cancer.org.au/assets/pdf/inequalities-in-cancer-outcomes#_ga=2.24233978.1745316963.1634631754-291144730.1632722355
3. Tervonen HE, Aranda S, Roder D, Walton R, Baker D, You H, et al. Differences in impact of aboriginal and torres strait islander status on cancer stage and survival by level of socio-economic disadvantage and remoteness of residence-a population-based cohort study in australia. Cancer Epidemiology. 2016;41:132-8.
4. Afshar N, English DR, Blakely T, Thursfield V, Farrugia H, Giles GG, Milne RL. Differences in cancer survival by area-level socio-economic disadvantage: A population-based study using cancer registry data. PLoS One. 2020 Jan 30;15(1):e0228551. doi: 10.1371/journal.pone.0228551. PMID: 31999795; PMCID: PMC6992207.
5. Bergin RJ, Whitfield K, White V, Milne RL, Emery JD, Boltong A, et al. Optimal care pathways: A national policy to improve quality of cancer care and address inequalities in cancer outcomes. Journal of Cancer Policy. 2020;25:100245.
6. Tervonen HE, Purdie S, Creighton N. Using data linkage to enhance the reporting of cancer outcomes of aboriginal and torres strait islander people in nsw, australia. BMC Medical Research Methodology. 2019;19(1).
7. Hoving JL, Monaco AD, Macfarlane E, Fritschi L, Benke G, McKenzie D, et al. Methodological issues in linking study participants to australian cancer registries using different methods: Lessons from a cohort study. Australian and New Zealand Journal of Public Health. 2005;29(4):378-82.
8. Clinical Oncology Society of Australia (COSA) Patient Reported Outcomes Working Group; Koczwara B, Bonnamy J, Briggs P, Brown B, Butow PN, Chan RJ, Cohn RJ, Girgis A, Jefford M, Jl Joske D, Licqurish S, Mackay G, Saunders CM, Webber K. Patient-reported outcomes and personalised cancer care. Med J Aust. 2021 May;214(9):406-408.e1. doi: 10.5694/mja2.50893. Epub 2020 Dec 13. PMID: 34046907.

Supplementary Materials

Characteristic	Question
1. Data Custodian	The Data Custodian is the Agency who holds overall accountability and responsibility for the dataset
2. Public research	For research arising from this database, is there a list of projects and/or findings from these studies available to the public?
3. Opt in/out	Do the people represented in these data have the opportunity to opt-in and/or opt out of providing their data?
4. Consultation	Is there a formal consultation process with researchers, government, health professionals and health advocates to determine the way in which data is collected, stored and accessed?
5. Data Security	Is the data collection managed under the "Five Safe Framework" (https://www.abs.gov.au/ausstats/abs@.nsf/Latestproducts/1160.0Main%20Features4Aug%202017#FIVESAFES)?"
6. Data governance	Is there formal written documentation in place to ensure data confidentiality, storage safety and maintenance of data quality? If so, what are they?
7. Routine transfer to national linked database	Are the data in this dataset routinely transferred to a national linked database?
8. Routine transfer to state database	Are the data in this dataset routinely transferred to any state-based linked datasets?
9. Routine transfer to national database	Are the data in this dataset routinely transferred to national dataset, which combines state/territory data and deduplicates records?
10. Unique identifier	Is there a nationally recognised unique identifier attached to each record in the dataset?
11. File formats	What are the standard formats that are provided for data extracts? (e.g., comma delineated)
12. Timeliness of transfer	If answer to 7 or 8 is yes, when is the data transferred to the national dataset? (e.g., immediately/6months/12months after notification of the individual record is received)
13. Enduring link	If answer to 7 or 8 is yes, is it an enduring link that is updated regularly?
14. Method of data transfer	If answer to 7 or 8 is yes, how is the transfer completed? (e.g., via high speed, secure portal, single transfer of password protected file, or other)
15. Management legislation/policy	Is there legislation/policy guiding the management of this dataset? If yes, what it is?
16. Plain language version	Is there a plain language version of this legislation that conveys its practical implications?
17. Legislation is common to other jurisdictions	Are these common to other jurisdictions?
18. Custodian transfer	If yes to question 7 or 8, when linkage occurs does the original custodian transfer data ownership to a single national linked-data custodian?
19. Access to de-identified information	What is the process for an external user to access de-identified (ie. no names, address, date of birth) unit record data from this dataset? (e.g., single/multiple access requests/ethics approvals, trusted user status, or no access for external users)

Supplementary Materials

Characteristic	Question
20. Access to de-identified information from Linked data	If the answer to 7 or 8 is yes, what is the process for an external user to access de-identified (ie. no names, address, date of birth) unit record data from the linked dataset? (e.g., single/multiple access requests/ethics approvals, trusted user status, or no access for external users)
21. Use for health system monitoring and improvements	Is the data currently being used to monitor and inform improvements to the health system
22. Use of linked data for health system monitoring and improvements	If the answer to 7 or 8 is yes, is the linked data currently being used to monitor and inform improvements to the health system
23. National Minimum Dataset	Is there a National Minimum Dataset for this data collection? If so, is the data collected in this dataset consistent with the National minimum dataset?
24. National Minimum Dataset for linked data	If the answer to 7 or 8 is yes, is there a standard protocol for the collection and the storage of data? If so, what is it?
25. Data cleaning process	Are there formal guidelines available to external users regarding the process of data cleaning and reporting and resolving of errors?
26. Data quality monitoring	Is there a formal guideline available to external users regarding the data quality monitoring for the data collection?
27. Data cleaning process – linked data	If the answer to 7 or 8 is yes, are there formal guidelines available to external users of the linked dataset regarding the process of data cleaning and reporting and resolving of errors?
28. Data quality monitoring – linked data	If the answer to 7 or 8 is yes, Is there a formal guideline available to external users of the linked data regarding the data quality monitoring for the data collection?

Questions in grey shaded cells are only relevant where data collection contributes to a linked dataset.

Cancer Council

cancer.org.au