Australian Research Data Commons

HeSANDA Development Priorities

Research community consultation report

Version 1.1

February 2021

Document version history

Version	Release date	Description
1.0	December 2020	Initial release of document
1.1	February 2021	● Addition of ‘Abbreviations’ section and footnotes ● Addition of ‘Next Steps’ section ● Addition of research community consultation report as Appendix A ● Minor typographical corrections and word choice revisions

Abbreviations

ACTA	Australian Clinical Trials Alliance
AHRA	Australian Health Research Alliance
AIHW	Australian Institute of Health and Welfare
ANZCTR	Australian New Zealand Clinical Trials Registry
ARDC	Australian Research Data Commons
CDM	Common data model
CRF	Case report form
DMP	Data management plan
DOI	Digital object identifier
EOI	Expression of interest
HeSANDA	Health Studies Australian National Data Asset
HREA	Human research ethics application
HREC	Human research ethics committee
IPD	Individual participant-level data
NHMRC	National Health and Medical Research Council
PHRN	Population Health Research Network
PICO	Population; Intervention; Comparison; Outcome
PISCF	Participant informant statement and consent form
TGA	Therapeutic Goods Administration
TRE	Trusted research environment

Executive Summary

The Australian Research Data Commons (ARDC) is partnering with the health research community to build a distributed national data asset from the outputs of health studies. National data assets increase the value of data collected by health studies by enabling deeper insights into our health and powering new research. The Health Studies Australian National Data Asset (HeSANDA) initiative aims to support health data sharing and secondary use in a way that brings value to the research community; increases the impact of research; maximises the investment in health research; and provides health, economic and social benefits to Australia's population.

HeSANDA development is incremental. The initiative is building consensus within the health research community and key stakeholders about the purpose, requirements, and practice guidelines for data sharing and secondary use. Initially, HeSANDA will assist research organisations in aligning their existing infrastructure to a coordinated national model for data sharing and secondary use.

Preliminary consultations with the National Health and Medical Research Council (NHMRC), Research Australia, Cochrane Australia, Australian Health Research Alliance (AHRA), Australian New Zealand Clinical Trials Registry (ANZCTR), Australian Clinical Trials Alliance (ACTA), and the Population Health Research Network (PHRN) identified broad aspirations and scope for the HeSANDA initiative. These organisations were invited to form an advisory committee to guide ARDC's approach for HeSANDA. The advisory committee recommended an initial proof-of-concept focusing on investigator-initiated clinical trials data sharing. This decision leverages the relative maturity of the clinical trials data community within Australian health research to deliver early value. The advisory committee identified the key stakeholder groups as researchers, research participants, research institutions/organisations, funders, and associated government bodies and regulators.

To assess the broader research community's response to the goals of the initiative and to build consensus on the initial scope of HeSANDA (including research purpose, value proposition, data requirements, and feasibility), ARDC ran a series of open consultations with the clinical trials and health research community. These consultations established in-principle support for HeSANDA and a set of principles to guide infrastructure development according to the needs of researchers. The consultation identified that:

- There is research community support for HeSANDA's vision to optimise research practices and thus maximise the public good by developing a coordinated national approach to sharing data from clinical trials research
- There is significant variability in data sharing practices and uncertainty regarding the best ethical and governance practices which impedes researchers wanting to engage in data sharing
- This variability and uncertainty impact the efficiency of research by increasing resource costs which present barriers for researchers
- HeSANDA can play an important enabling role in supporting Australian health research efficiency, impact, and translation by addressing these issues at the national level, establishing data sharing capability, and promulgating standards that will make data sharing and secondary use more efficient

The consultations indicated a clear priority for the sharing of individual participant-level data (IPD) from clinical trials to support research efficiency, impact, and translation. Recommendations for investment into three areas to enable this development were identified:

1. Coordinated national network of data services
2. Centralised data discovery and access tools, and
3. Coherent data practices.

The scope of the initiative and the requirements of the HeSANDA infrastructure will be further refined by future engagement with the key stakeholder groups.

Research community priorities

A series of consultation workshops established HeSANDA's core business requirements, value proposition, and high-level feasibility from the perspective of the clinical trials research community. Structured consultations addressed four themes:

- Theme A: "Research purpose" (users, use cases, value proposition, priorities)
- Theme B: "Data content & quality requirements" (information and data needs)
- Theme C: "Existing data standards & practices" (landscape analysis)
- Theme D: "Barriers, systems, & enablers" (feasibility and incentives)

The Australian Institute of Health and Welfare (AIHW) assisted ARDC to design and deliver the consultation series guided by foundational steps in their 'data development' process[1]. ARDC extended an open invitation to participate in the consultation workshops to researchers, institutions, organisations, and policy makers involved in clinical trials research. ARDC appointed an editorial team of eight health research and infrastructure professionals drawn from external organisations to facilitate the consultation workshops, analyse the feedback, and draft the findings presented below and in the detailed consultation report provided in Appendix A.

Summary of consultation outcomes

The consultation process validated and refined HeSANDA's aspirational requirements from research community members across the research journey, from the early stages of research ideation of a clinical trial through to secondary use of data collected by that trial. The consultation also reviewed existing standards and practices used in clinical trials research, and investigated the data content and quality characteristics required to support HeSANDA's core value streams.

Consultation identified seven research areas where shared clinical trials data would return significant value to the research community, with items 1-3 (below) being of primary importance and underpinning the other areas:

1. **Meta-analysis and systematic review**
2. **Replication, reproducibility, &/or peer review**
3. **Secondary research projects and analyses**
4. Policy development
5. New study design
6. Health technology assessment
7. Clinical guideline development

The primary information required to enable these kinds of research is individual participant-level data (IPD) from clinical trials. The minimum information required to contextualise IPD were identified as study protocols, data descriptions, and data quality statements. Study summary and ethics information are required to support data discovery and access.

The consultation process identified gaps in research data practices and feasibility considerations relevant to the development of data sharing practices and infrastructure. These are discussed in detail in the

[1] Australian Institute of Health and Welfare (AIHW) 2007. A guide to data development. AIHW Cat. no. HWI 94. Canberra: AIHW.

consultation report (Appendix A) and summarised in the Principles section below. In broad terms, the consultation process established that:

1. Australian researchers represented in the consultations are enthusiastic about the HeSANDA initiative: both data producers (including the clinical trialists and participants) and secondary data users expressed support for HeSANDA's vision. In particular, consultation participants emphasised HeSANDA's value as a vehicle for sharing and accessing high-quality IPD and associated metadata. Researchers reported strong incentives to engage with HeSANDA that pertained not only to enhancing their own professional activity, but with **maximising the public good** that could be achieved by developing a more efficient approach to data management and sharing.
2. There is **currently a lack of standardisation** in the management of clinical trial data both with respect to the data systems and practices that different research groups use, as well as the implementation of data governance by different custodians and jurisdictions. Indeed, there is uncertainty around the ethical and consent requirements regarding secondary data use that presents a barrier to researchers wishing to engage in data sharing. HeSANDA will need to take an active role in addressing not only issues of data management practices and data standards but also data governance in order to support successful data sharing.
3. Despite community enthusiasm for data sharing and secondary use, **the high costs of standardising and sharing data was reported as a barrier.** Similarly, for secondary data users, the costs of finding and securing the use of existing data are currently also very high. These include the resource costs associated with searching for suitable data when the publicly-available metadata are often insufficient to determine whether a specific dataset is fit-for-purpose for the secondary user.
4. **HeSANDA can play an important enabling role in supporting Australian health research efficiency, impact, and translation by addressing these issues at the national level** and promulgating standards that will make data sharing and secondary use more efficient. This includes, but is not limited to, adopting a leading role in standardising both data governance frameworks (including participant consent) and also the conventions and mechanisms that are adopted nationally for data sharing and secondary use.

Principles

Based on the feedback obtained via the consultation process, the editorial team derived a list of principles for HeSANDA which reflect the needs and perspective of the clinical trials research community who participated in the initial data development workshops. The purpose of these principles is to provide direction that will assist HeSANDA to achieve its vision and mission to promote data sharing and secondary use in Australia. The workshop feedback informing each principle is listed in the 'source' column.

Principle	Statement, rationale, further information	Source
Purpose		
[1] The capabilities delivered by HeSANDA must be informed by the core value proposition	HeSANDA will enable the national infrastructure required to support the sharing and secondary use of health research data in order to improve research efficiency, reduce cost, and increase research impact.	Theme A

[2] The core research purpose of HeSANDA is to support research with a translational focus	HeSANDA will support access to the information and outputs from clinical trials necessary for: • meta-analysis and systematic review • replication, reproducibility, &/or peer review • secondary research projects and analyses to facilitate the translation of research into clinical guideline and policy development, health technology assessment, and the development of new research.	Theme A
[3] HeSANDA will facilitate the sharing of a range of clinical trial information	To meet the needs of data producers and secondary users, HeSANDA will support the sharing of a variety of different types of information associated with clinical trials, with an emphasis on individual participant-level data, study protocol metadata, and cohort summary data.	Theme A, C
[4] HeSANDA will maximise the discoverability of the clinical trial information	The information available through HeSANDA must be organised in a way that supports efficient search and discovery of clinical trial information (e.g. using the PICO[2] framework).	Theme A, B
[5] HeSANDA will improve the efficiency and reliability of access to clinical trial data for secondary research	Trial information is currently siloed, predominantly stored on institutional servers, and often accessible to secondary researchers only via direct contact with individual trialists. However, there is a clear community enthusiasm for making this information accessible via more standardised and potentially centralised mechanisms to achieve optimal research efficiencies.	Theme C
[6] HeSANDA will reduce the barriers to data sharing for clinical trialists	In order to reduce resource costs and facilitate development of the data asset, HeSANDA must align data sharing with the existing research practices of trialists.	Theme D
Data content & quality		
[7] HeSANDA will promote minimum reporting & data sharing requirements for clinical trials	The implementation of minimum requirements maximises the utility of clinical trials data. Minimum data sharing requirements should include IPD and the study information that contextualises it (i.e. study protocol; data descriptions; data quality statements). Research and data descriptions to support Principle [4] must also be included. Enabling coherent data practices throughout the research journey can support these requirements.	Theme A, B, C, D
[8] HeSANDA will support the current variety of IPD data standards but will encourage pathways to the adoption of stakeholder-endorsed data standards	Currently, there is a wide variation in the data formats used to collect, enter, and analyse new data. As such, HeSANDA will need to facilitate the sharing of different data formats for IPD and metadata. However, HeSANDA should support the adoption of standardised data platforms and data standards for storing data and recording metadata (e.g. data dictionaries).	Theme C, D

[2] Richardson, S., Wilson, M. C., Nishikawa, J., & Hayward, R. S. (1995). The well-built clinical question: a key to evidence-based decisions. ACP journal club, 123(3), A12-13.

[9] Data quality statements will underpin the utility of HeSANDA's content	In order to provide confidence in the data asset, data quality should be represented for each data collection.	Theme A, B, C
Data governance		
[10] HeSANDA should promote common approaches to data sharing and re-use by clinical trials researchers	Researchers encounter resource and efficiency issues due to the lack of clear guidance on how to implement the data sharing policies of funders, publishers, and other stakeholders. The development of agreed protocols and procedures will improve the feasibility for data sharing to become standard research practice.	Theme A, C, D
[11] HeSANDA should promote common approaches to participant consent requirements for data sharing and re-use	Researchers agree on the fundamental importance of consent and community support for research practices such as data sharing. But as with the previous principle, they require guidance on how best to implement open science policies as they relate to participant consent. Developing a coordinated national approach to meet consent requirements will not only improve the feasibility of data sharing but, most importantly, address the concerns and mitigate risk around the sharing of sensitive data.	Theme A, D
[12] HeSANDA should promote best practice guidelines for the handling and sharing of sensitive data	To complement the principles of common approaches to policy interpretation and application (above), researchers will benefit from guidance on specific data handling issues such as data de-identification, security, etc.	Theme A, C, D
[13] HeSANDA should be considerate of the labour cost to clinical trialists to facilitate access to data	The above principles seek to improve the efficiency of data sharing (either directly or indirectly), thereby reducing costs and improving feasibility. However, these improvements cannot entirely remove the labour cost of data sharing that is not consistently supported at the funder or institutional levels at present. Recognition of these costs within data sharing policy and infrastructure is fundamental to supporting the research community.	Theme A, D
Stakeholder coordination		
[14] HeSANDA should align its activities with existing structures and initiatives that support the national harmonisation of clinical trial activities	For example, currently clinical trial researchers are required to enter common data regarding their trial in the human research ethics application (HREA) form, trial registration (e.g. via ANZCTR), and, where applicable, to the Therapeutic Goods Administration (TGA). To reduce administrative burden for researchers, HeSANDA will link to these and other existing structures to support better knowledge discovery and easier meta-analysis.	Theme A, B, C

[15] HeSANDA should attempt a nationally coordinated approach to address its data governance aspirations and principles	Issues of data sharing & governance impact multiple stakeholder groups, from research participants and researchers through to funders, institutions, and ethics committees. The research community desires cooperation and coordination between these groups to address their common interests.	Theme A, D
[16] HeSANDA should leverage existing investment in data sharing infrastructure where possible	Researchers are required or incentivised to utilise existing data management and sharing infrastructure provided by their organisations. HeSANDA should engage with research organisations in order to develop strategies to avoid unnecessary duplication of effort and to maximise existing infrastructure investments.	Theme C, D

Priorities for infrastructure development

Based on the consultation series, the editorial team identified gaps at various stages of the research journey that reduced the feasibility of data sharing and secondary use. These gaps constitute three priority areas for investment to enable data sharing:

1. A set of **coordinated data services** that:
    - Facilitate access to IPD for secondary use
    - Facilitate access to study summary information, protocols, data dictionaries, data quality statements, and ethics information to enable research discovery and secondary use of data
    - Support common data and metadata standards
    - Supply standardised descriptions to central discovery services (not held elsewhere)
    - Provide access to data according to a common governance framework
    - Support centralised data request and access processes
    - Provide tools for researchers to efficiently meet the above requirements
2. A set of **federation services** that integrate the coordinated data services to enable:
    - Research and data discovery
    - A streamlined data request process
    - Efficient data access
3. A set of stakeholder-endorsed **coherent data practices** for:
    - Research data and metadata standards
    - Standardising compliance with ethics and participant consent requirements
    - Data governance
    - Data request and access processes
    - Tools to facilitate data standardisation and compliance

These services and practices should adhere to two key requirements:

- Data sharing should support the interests of data producers and secondary users, research participants and the general public, research institutions and organisations, funding agencies and policy makers. The investment into infrastructure development should obtain the support and endorsement of these groups.
- While the potential scope for HeSANDA is boundless, identification of key types of data, evaluation of data availability, and current clinical trial policies / procedures can inform a phased rollout strategy. To be feasible, HeSANDA should be implemented in stages.

Next steps

The consultations established in-principle support for HeSANDA from the research community and identified the principles and priorities of the clinical trials research community for a national data asset. The scope of the HeSANDA initiative will be further refined by future engagement with the key stakeholder groups identified by the HeSANDA advisory committee. These requirements will feed into the HeSANDA Infrastructure Development stream of activity to guide the infrastructure design and development. Details on the scheduling and coordination of these activities can be found on the HeSANDA website.

Appendix A

Research community consultation report

Edited by Lisa Askie, Tiffany Boughtwood, Douglas Boyle, Luke Connelly, Anitha Kannan, Manuel Nielsen, Claire Vajdic, & Melina Willson

Overview

The aim of the research community consultation series was to build consensus around the purpose, contents, and other requirements for a national health data asset to be established under the Health Studies Australian National Data Asset (HeSANDA) initiative by the Australian Research Data Commons (ARDC). The focus of the consultations was on the sharing and secondary use of data from investigator-initiated clinical trials. The consultation workshops were designed and delivered with the assistance of the Australian Institute of Health and Welfare (AIHW) and guided by foundational steps in their data development process[3] which address: [a] business context and information need identification; [b] feasibility analysis; [c] consultation and collaboration; and [d] identifying data for development. An editorial team was recruited to support the planning and facilitation of the consultation workshops and to document the feedback and outcomes of the consultation series.

Consultation design

Participants

An open invitation to participate in the consultation series was extended to researchers, institutions, organisations, and policy makers involved in clinical trials research. Consultation workshops were announced via the ARDC website and newsletter; targeted emails to the National Health and Medical Research Council (NHMRC); Research Australia; Cochrane Australia; Australian Health Research Alliance (AHRA); Australian New Zealand Clinical Trials Registry (ANZCTR); Australian Clinical Trials Alliance (ACTA); Population Health Research Network (PHRN); Association of Australian Medical Research Institutes (AAMRI); outreach to respondents to a prior call for expressions of interest (EOI) to participate; and via ARDC staff on an ad hoc basis.

The invitation to the workshop series received a total of 93 registrations with the majority of registrants identifying as employees of Australian universities or medical research institutes. At the time of registration, registrants were asked “What is your involvement with clinical trials research?” and selected from six options. The frequency of responses is shown in Table A.1 below (N.B. 31 participants selected more than one option). Registrants who responded “Other” were asked for more details, however the details provided (e.g. “biostatistician”, “database manager”, etc) suggested that they were congruent with one of the five other categories (i.e. researcher, research management &/or support, etc).

[3] Australian Institute of Health and Welfare (AIHW) 2007. A guide to data development. AIHW Cat. no. HWI 94. Canberra: AIHW.

Table A.1. Registrants' roles in clinical trials research

Researcher	45
Research management &/or support	65
Policy maker	9
Funder	1
Participant	4
Other	9

Workshops & feedback options

Facilitated online workshops were conducted with agenda topics and key questions addressing specific consultation themes (see below) to enable both structured and unstructured feedback. The workshops consisted of a presentation followed by breakout groups. Moderated Q&A sessions and live polling were used where appropriate.

Participants and roles:

- ARDC program manager - co-facilitator guided discussion on HeSANDA objectives
- AIHW representative - co-facilitator on technical aspects of data development
- Editorial team - facilitated, documented, and analysed survey responses and breakout room discussions
- Invited speakers - research professionals giving presentations on their insights and perspectives on workshop themes
- Workshop participants - members of the clinical trials and health research community who self-nominated to attend one or more workshops to contribute to HeSANDA requirements

Structure: Four 90-minute workshops were conducted (one every 3 weeks) from August to October 2020. A maximum of 45 people attended per workshop, and small breakout group sessions included up to 12 participants per group.

Supporting documentation: Reference documents which introduced the topic and focus questions were circulated to participants one week prior to each workshop. Recordings of the workshop presentations were published within 48 hours and registrants were notified of their availability.

Feedback options

Feedback was sought from registrants via multiple mechanisms depending upon the theme and nature of information to be captured:

1. Workshop participation
2. Open-ended written feedback (i.e. participants could provide a written submission to indicate their interests, opinions, and/or use cases)
3. Structured feedback (via structured questionnaires on specific topics and focus questions)
4. Participation via voting mid-workshop (Mentimeter)

Attendees had the choice of providing feedback by as many mechanisms as they preferred and were advised that they would not be identified in the reporting of their feedback. Registrants who could not attend the workshops were invited to provide feedback via options 2 and 3 where available.

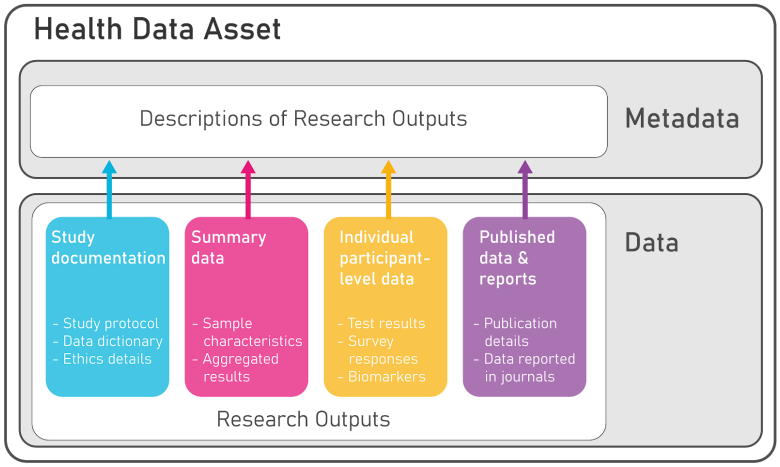
Consultation themes

Concepts & definitions

The concepts and terminology used to discuss research data vary between sectors, organisations, and contexts. This section provides a brief discussion on the use of some of the concepts and terminology within the consultation series and this report.

Research outputs, data, and metadata: The clinical trials research community predominantly uses the term 'data' to refer to individual participant-level data (IPD), but can also use it to refer to study cohort characteristics, published results, and other information about a trial. Similarly, they use the term 'metadata' non-specifically to refer to anything other than IPD (e.g. summary data, study documents and data dictionaries, etc). However, within the context of a national data asset, the 'research outputs' of a clinical trial (i.e. any information, IPD/data, and/or documentation generated by or about a clinical trial) constitute the *data* held by the data asset, while the descriptions of these research outputs are the *metadata*. These concepts can be illustrated as follows:

Figure A.1. Illustration of the relationship of health research outputs to 'data' and 'metadata' in the proposed HeSANDA data asset



In this report, the term 'research output' has been used in place of 'data'. Similarly, for compound terms such as 'data sharing', 'data standards', etc, 'data' should be interpreted as 'research output'.

Data standards and practices: The AIHW defines data standards as a way to "describe the agreed meaning and acceptable representation of data for use within a defined context" that "enable data from different sources, organizations or systems to be exchanged and compared in a meaningful way"[4]. The term also has a specific technical implementation within their operational context. However, for the purposes of the HeSANDA consultations, 'data standards' was used in a general sense to refer to any commonly used practices or conventions for the documentation, coding, and handling of research outputs.

[4] Page 13, Australian Institute of Health and Welfare (AIHW) 2007. A guide to data development. AIHW Cat. no. HWI 94. Canberra: AIHW.

Data quality: 'Data quality' and 'data quality statements' have specific technical implementations within some operational contexts. For the purposes of the HeSANDA consultations, the terms were used in a general sense to refer to the extent to which research outputs (but most often IPD) are fit for secondary use. Since secondary use extends beyond the primary purpose of data collection, an a priori determination of fitness for use is challenging. This topic is discussed further in the Analysis of Feedback section below.

Workshop themes & design

The consultation workshops were based on four themes:

- **Theme A**: **Research purpose** - How do researchers wish to use shared data? What kinds of information are needed for these uses? Who would use shared data? How should these uses and purposes be prioritised?
- **Theme B**: **Data content & quality requirements** - Based on the purpose and uses identified in Theme A, what data types and other primary research outputs need to be included in the data asset, and what are the quality requirements for these data?
- **Theme C**: **Existing data standards & practices** - Based on the data types, research outputs, and quality requirements identified in Theme B, what data standards and practices exist or would need to be developed?
- **Theme D**: **Barriers, systems, & enablers** - What are the issues (e.g. data governance, access arrangements, IT & infrastructure) that must be addressed in order for data sharing to be implemented successfully?

The themes were guided by AIHW's data development process and adapted to suit HeSANDA's requirements. Each successive theme built upon previous themes to develop a broad understanding of participant requirements and to confirm and refine feedback about previous themes. A summary of the workshop design and attendance is reported in Table A.2.

Editorial team

The editorial team was recruited via an open EOI call. Applicants were assessed on the following criteria:

1. Experience in any of the following:
    - research health studies projects
    - research health studies data infrastructure
    - health studies meta analysis, guideline development or data linkage
    - health data standards development
2. Experience in requirements analysis or consensus documentation

The editorial team assisted with the planning of the workshops and reference documents, facilitated breakout sessions during the workshops, analysed all feedback (including written submissions, survey/polling data, and workshop discussion), and drafted the research community priorities reported in the main body of this document and the consultation report presented in this appendix. The editorial team members were:

- Lisa Askie (University of Sydney)
- Tiffany Boughtwood (Australian Genomics, Murdoch Children's Research Institute)
- Douglas Boyle (University of Melbourne)
- Luke Connelly (University of Queensland)
- Anitha Kannan (Monash University)
- Manuel Nielsen (South Eastern Sydney Local Health District)
- Claire Vajdic (University of New South Wales)
- Melina Willson (University of Sydney)

Table A.2. Workshop themes & design

Workshop title	**Theme A: Research purpose & value proposition**	**Theme B: Data content & quality requirements**	**Theme C: Existing data standards & practices**	**Theme D: Barriers, systems, & enablers**
Date	11 August 2020	1 September 2020	22 September 2020	13 October 2020
Aim	To establish the research purpose and value proposition of HeSANDA by identifying: ● who might use the data asset; ● how they might use shared data; ● the types of information needed for these use cases; ● prioritisation of these activities.	To determine the data content and quality of HeSANDA required to meet the needs identified in Theme A by establishing: ● which data is to be included; ● what research outputs should be included; ● which data should form the minimum requirements for reuse; ● what should be the quality requirements of the included data.	To identify the existing data standards and practices within the Australian research community in order to assess: ● which data standards are already in use; ● which data standards are preferred; ● what practices exist; ● what practices need to be developed; ● the alignment of existing standards and practices with a draft HeSANDA information scope based on feedback about Themes A and B.	To identify the practical and implementation issues that impact researchers' ability and willingness to engage in data sharing: ● incentives to share data; ● barriers to data sharing; ● the relative impact of incentives and barriers; ● systems and enablers that can address these issues.
Topics	● Research purpose ● User types ● Information types ● Use cases ● Value proposition	● Existing data sharing platforms ● Minimum information requirements ● Data content ● Metadata ● Data quality ● Data standards	● Draft information scope ● Current and preferred data practices	● Costs & feasibility ● Ethics issues ● Data governance ● Incentives for data producers
Attendance & feedback	● 45 workshop attendees ● 41 participants in breakout room discussions ● 10 respondents provided structured feedback via an online survey ● 1 respondent provided unstructured feedback via written submission	● 30 workshop attendees ● 30 participants in breakout room discussions ● 8 respondents provided structured feedback via an online survey ● 6 respondent provided unstructured feedback via written submission	● 24 workshop attendees ● 24 participants in breakout room discussions ● Structured feedback via an online survey was received from: ○ 23 clinical trialists/data producers ○ 17 secondary data users	● 27 workshop attendees ● 27 participants in breakout room discussions

The AIHW reviewed and provided additional feedback on the draft report, and the final version of this document was prepared by ARDC with the endorsement of the editorial team and AIHW support.

Analysis of feedback

Each workshop was structured around one of four distinct themes (see Table A.2), however, participants often provided feedback about specific topics across multiple workshops. As such, the following analysis and reporting of the feedback received is grouped by topic rather than the workshop themes in order to assist the reader to efficiently navigate the content. The Summary of Feedback section which follows then harmonises the topics and reports the feedback received for each main theme.

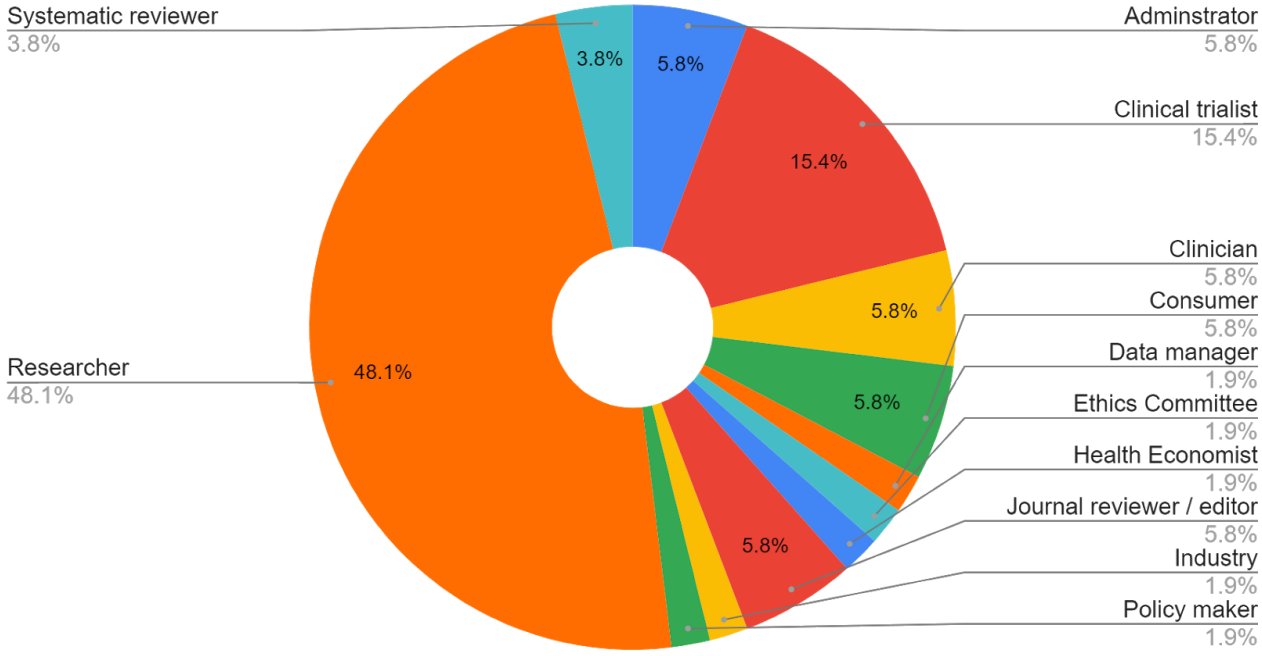
Research purpose of HeSANDA

To gain insight into the research purpose, breakout group discussions were used during the first workshop to generate user stories. Respondents were asked to complete the following statement: "*As a [user type], I want [information type], so that [use case]*". A structured survey was also offered to capture this information from registrants who were unable to attend or for attendees who wished to provide further detailed feedback. An analysis of the responses is as follows:

User types

The survey respondents and workshop attendees were principally researchers (48%) and clinical trialists (15%) with a small representation from other stakeholders such as consumers, journal editors, policy makers and those with process-related roles such as ethics and data management. It should be noted that many respondents have more than one role and some of the use-cases given by respondents reflected their wider interests and responsibilities. The participant breakdown is provided below, as Figure A.2.

Figure A.2. User types



The number of individuals who participated meant that feedback on use cases in some categories of importance (such as systematic reviewers) was limited. For this workshop, the small number of technical implementation specialists was of less significance as we wished to determine top-level research aspirations of the HeSANDA program with the more granular technical specifications to be addressed in later consultations.

Research uses (use cases)

Workshop participants were asked to consider their aspirations with regard to information from or about clinical trials from a variety of different contexts (e.g. as a researcher or as a policy maker). The participants were asked to explain what they wanted, for what reason, and to provide any special requirements in relation to this aspiration. The process was not designed to identify all known aspirations of stakeholders but to identify aspirations that are partially or wholly unmet. Seven breakout sessions were held to identify aspirations across different research areas:

- Meta-analysis and systematic review
- Policy development
- New study design
- Replication, reproducibility, &/or peer review
- Secondary research projects and analyses
- Health technology assessment
- Clinical guideline development

The workshop gave broad consensus around the themes above with a strong focus in three areas:

1. **Secondary research projects and analyses**
2. **Meta-analysis and systematic review**
3. **Replication, reproducibility, or peer review**

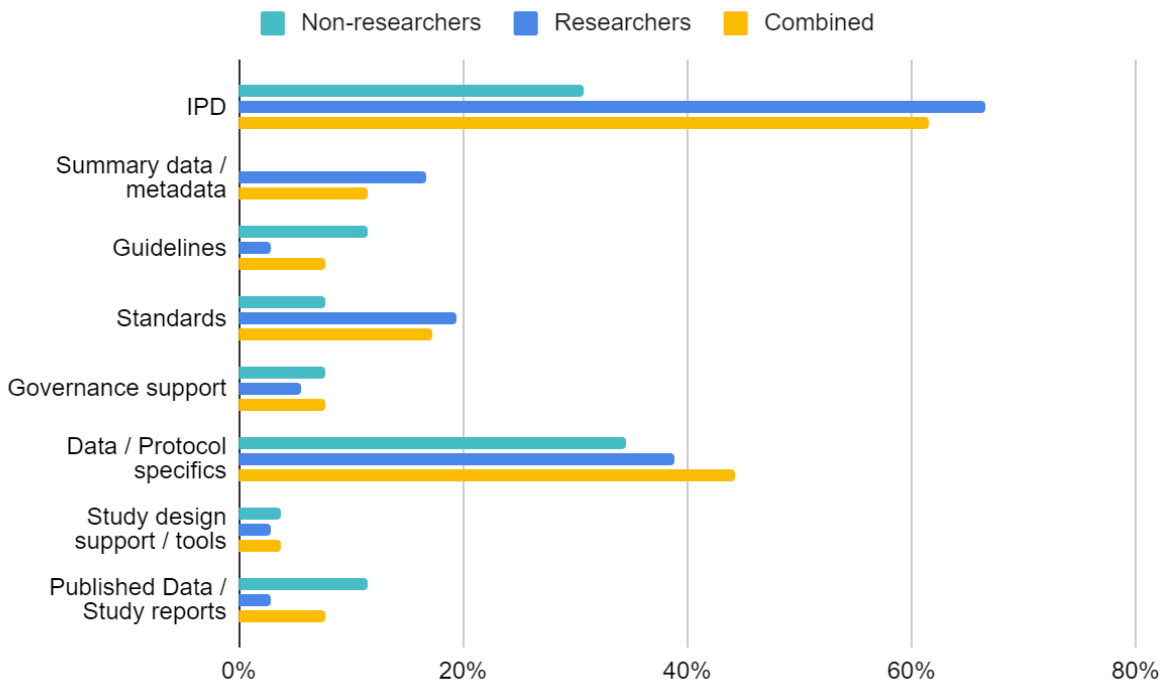
It is noted that standards that can be applied to new study design will help enable these three areas of consensus. Policy development, health technology assessment and clinical guideline development can be enabled by the development of infrastructure to support the focal areas listed above.

Information types

A range of information types were specified as being required to facilitate the identified research uses. The majority of respondents, independent of user type, indicated they wanted IPD and specific information about the study data/protocol. Many specified the need for research outputs to be presented in a standardised fashion, and summary data/metadata availability. Other proposed HeSANDA information types included information about published data, study reports, and study guidelines. During these discussions, participants identified resources such as study design support/tools and governance support that, while not being research outputs, could support data sharing activities.

These information requirements are represented in the figure below, noting that these are indicative proportions only, derived from thematic analysis of respondents' comments/written submissions. The figure indicates the respective requirements of 'researchers' (i.e. respondents who identified as researchers, clinical trialists, systematic reviewers, and health economists) and 'non researchers' (i.e. all other reported user types) in addition to the overall group response.

Figure A.3. Required information types



Both subgroups expressed similar prioritisation of information types which seem to reflect the existing challenges associated with obtaining IPD and also reflect a common desire for information such as study protocols and other summary information and documentation to be more easily available - and for these to be standardised wherever possible. Specific information about study protocol, definitions of the variables collected, data standards, and summary data were also identified as important attributes of any collections. Notable comments and themes were:

- **Individual participant-level data**: respondents expressed an interest in accessing IPD for the purposes of secondary analysis for new studies - to explore potential novel correlations, suggest new hypotheses, validate correlations suggested by smaller studies, undertake longitudinal studies and provide evidence for prospective trials. Others were keen to see their results (including IPD and study protocols) re-used by others for greater research impact, and the potential for future acknowledgement and authorship.
- **Data/protocol specifics**: access to detailed information about the study design and data capture was perceived as critical to the use of the data asset. Stated uses included the ability to judge the validity and applicability of an intervention to patient cohorts; to permit replication and merge data across studies; and to judge the quality of the study – and hence its IPD and research outcomes.
- **Other themes**: respondents often referred to the need for discoverability of clinical trials and the searchability of their research outputs, as well as the critical need to trust in the quality, and standardisation of these outputs. The importance of prospective application of standards to a study was emphasised, due to the burden of 'retro-fitting' standards to an existing dataset. Consent, data management, and data sharing plans were also highlighted frequently by respondents to assess the ethical aspects of research conduct and compliance, or to inform the design and approval of future studies.

Existing data sharing platforms

Workshop participants viewed demonstrations of data sharing platforms built to facilitate discovery of and access to clinical trial datasets. The platforms demonstrated were: Clinical Trial Data Request; Vivli; Figshare; and Yoda[5]. Participants were asked to reflect on their experience with the use of such platforms, or alternatively, what should be included in the minimum requirements for a data sharing platform to be of most use. Responses from data providers and secondary users indicated minimal use of platforms designed specifically to support data sharing (e.g. Vivli) because they were not considered fit-for-purpose. Participants reported using the following platforms and services when planning or performing meta-analyses:

- ANZCTR[6]
- Figshare
- Australian Data Archive[7]
- AARnet services[8]
- Australian institutional data repositories
- International data repositories

Some of these data sharing platforms are searchable but on the basis of limited information. As a result, significant interaction with the data provider and/or review of IPD is needed to determine whether study data can be re-used for a specific purpose.

Data providers reported routinely using digital platforms to submit ethics applications (including protocols, case report forms (CRF), questionnaires, participant information and consent forms (PISCF), etc), amendments, annual reports and adverse event reports about their trials. These platforms thus house much of the information required for secondary use of data, but they can only be accessed by the study's primary investigator, their nominees, human research ethics committee (HREC) and institutional governance officers.

Participants expressed a range of features that they would like to see in a data sharing platform. There was strong support for the integration of existing publicly available sources of research outputs so that all information was accessible in a single place. This was seen as important in terms of efficiency for data providers but would also ensure comprehensive coverage of research outputs and the inclusion of approved study documentation.

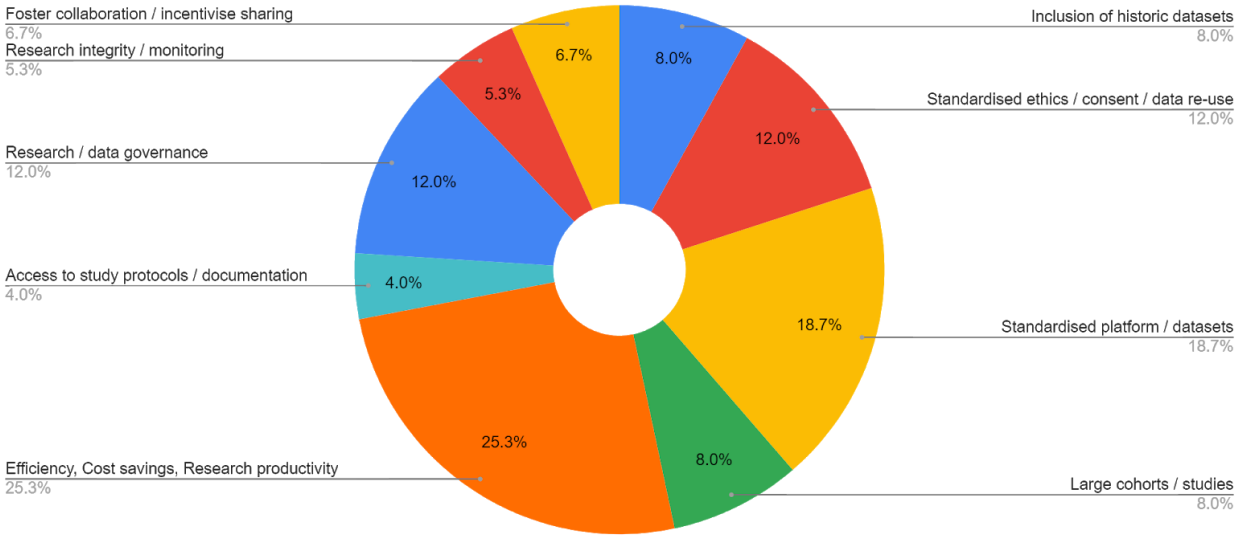
Value proposition for HeSANDA

Respondents were asked 'How would you get value from HeSANDA?'. Thematic analysis revealed nine broad categories of value or benefit from HeSANDA, with most respondents identifying more than one potential benefit from the project (on average, respondents nominated 1.4 benefit categories). Figure A.4 (below) provides a breakdown of respondents' views on the value proposition offered by the HeSANDA initiative.

HeSANDA's potential to improve research efficiency, productivity and yield cost savings was a common theme amongst workshop participants; with a quarter of respondents specifically commenting on the opportunity to reduce waste and duplication, save time, avoid duplicating data collection, and mitigate project risk.

[5] https://www.clinicalstudydatarequest.com/; https://vivli.org/; https://figshare.com/; https://yoda.yale.edu/
[6] http://www.anzctr.org.au/TrialSearch.aspx
[7] https://ada.edu.au/
[8] https://www.aarnet.edu.au/network-and-services/collaboration-services/

Figure A.4. HeSANDA Value Proposition



The benefit of a standardised platform and IPD datasets was another frequent theme in respondents' value propositions, noting the importance of reducing barriers/cost of data access, and providing a means to search across datasets. One important aspect of the latter theme was the view that dataset contents (e.g. variable names and definitions, units of measurement, and other IPD metadata) should be clearly documented so that potential data users could determine the likely usefulness of an archived dataset for their purposes without needing to access the IPD itself.

One respondent noted: "Having a searchable resource, with consistent metadata, would be a great advantage: the ability to search and find which groups of researchers are conducting work on specific topics … would form an important source of knowledge where, currently, researchers have to rely essentially on network-related information (e.g. professional contacts in the field)."

Respondents also cited the value of having access to historical datasets; the benefit of access to and building large cohorts; the opportunity to monitor research integrity/detect fraud; and to foster collaboration and incentivise sharing.

In summary, it was clear that participants viewed the value HeSANDA will create will be realised through the delivery of efficiencies gained through standardisation of research outputs and standardisation of process. These efficiencies are the most important component of the value proposition offered to stakeholders by the HeSANDA initiative.

Other comments

Beyond commenting on the research use and value proposition of HeSANDA, some respondents provided additional information about other potential functionalities or incipient opportunities for HeSANDA:

- Many respondents commented on the potential value of supporting standardised ethics or consent systems, to support data re-use, and research/data governance. This topic was explored further in a later workshop but the foundational nature of ethical research practice, and data use in line with

participant consent, meant respondents frequently referred to these issues throughout the consultation series.
- The potential to improve trial participant experience, improve privacy, or be a mechanism for reimbursement was mentioned by one respondent.
- Another respondent proposed training and credentialing for 'trusted' researchers who could be allowed to access research outputs via an expedited process.
- The possibility of linkage with other datasets, such as electronic medical records, was on the wish-list of some respondents; and one was particularly keen on a dedicated HREC to facilitate approval and access.
- Some respondents expressed scepticism about the feasibility of HeSANDA, or the possibility of the program's delivery. Some commented on the expense of preparing data for sharing/secondary use; doubted the return on investment of this data re-use; and noted the potential risk to ARDC in the HeSANDA program. Most of these criticisms were accompanied by assurance of the value of the initiative in principle and were thus concerned about the delivery rather than the goals.

Information needs

Consultation participants consistently identified IPD as the most valuable research output for inclusion in HeSANDA to support their identified research purposes. This priority was expressed repeatedly throughout the breakout session discussions on research purpose, value proposition, and information needs (see 'Research purpose of HeSANDA' above for an indication of the level of prioritisation). Participants also reported that the inclusion of summary data and other research outputs would be necessary for effective secondary use of IPD. The information priorities were:

- Study protocol
- Data dictionary
- Proof of ethical approval
- Proof of participant consent to data re-use

Participants also frequently reported that the access to and/or information about the following research outputs would greatly facilitate the re-use of IPD:

- Trial registration details
- Publications, preprints and abstracts
- IPD data quality statements
- CONSORT[9] statement and extensions
- Unpublished reports
- Lay description
- Contact details to request IPD re-use
- Data management plans
- Statistical analysis plans

Participants provided additional more granular feedback on specific information needs as follows:

Clinical trial summary data

The inclusion of study summary data in HeSANDA, specifically summary statistics on patient characteristics and the outcome variables, was frequently recommended on the basis that it would increase the efficiency of

[9] Consolidated Standards of Reporting Trials (http://www.consort-statement.org/consort-2010)

data re-use. It would allow for the rapid review of study characteristics and data quality necessary to determine study eligibility/suitability for re-use prior to requesting approval to access IPD.

Participants suggested that study summary data and the descriptions of research outputs should be standardised to facilitate efficient discovery of clinical trial information (e.g. using the PICO framework). Respondents reported that the minimum descriptive information required to support that is:piscf

Table A.3. Minimum descriptive information to support research discovery

Participant	Clinical diagnoses
	Age (group/band)
	Sex
	Race
	Comorbidities
Intervention	Specific intervention tested
	Category (domain): • pharmaceutical • psychological • device • other
	Drug class (if pharmaceutical intervention)
Outcomes	Care outcome
	Adverse events
Sample size	Recruitment (target and actual)

In terms of trial outcomes, secondary users of data gave weight to the ability to identify specific adverse events in addition to the pre-specified primary and other outcomes. Secondary users also expressed a strong need to access information about the achieved sample size, including for subgroups, the extent of missing IPD at the variable level, and the study publication status, as the ability to search for these features would greatly enhance the efficiency of data re-use.

It was recognised that there are efforts being made to standardise study summary information, most notably the implementation of research ethics systems (REGIS[10] in NSW, RGS in WA, ERM[11] in VIC, QLD, TAS), SPIRIT[12], CONSORT, and various clinical trial registries. For information about trial publications, preprints and abstracts, publishers and other organisations already maintain extensive catalogues and digital object identifiers (DOI) for these research outputs. Utilising these existing systems is highly desirable as it would reduce the burden and potential discrepancies between the same information being managed in multiple systems. Participants were highly supportive of a data asset that was interoperable with these existing systems and could aggregate the various information about trials that is spread across them.

[10] Research Ethics and Governance Information System (https://regis.health.nsw.gov.au/)
[11] Ethical Review Manager (https://au.forms.ethicalreviewmanager.com/)
[12] Standard Protocol Items: Recommendations for Interventional Trials (https://www.spirit-statement.org/title/)

Study documentation

Participants identified study protocols and data dictionaries as essential documents to support the secondary use of IPD. Ideally, the study protocol would follow industry best practice (e.g. CONSORT, SPIRIT) and include detailed information on the participant inclusion/exclusion criteria, the recruitment strategy, the randomisation, the blinding, the intervention(s), the CRF, any questionnaires or other measurement tools, and the PISCFs. Furthermore, the data dictionary would include data definitions, and the consent information would include any conditions related to data re-use, such as the research scope.

Participants also expressed a strong desire for standardised templates to help in the design of new studies. Suggested templates included:

- Protocol (e.g. CONSORT, SPIRIT)
- PISCF that includes ethically approved text about data re-use
- Data sharing agreement
- Data management plan
- Data monitoring plan
- Statistical analysis plan

Individual participant-level data

To maximise utility of IPD, participants recommended HeSANDA include a clear description as to whether the IPD corresponds to the entire study population, the study publication(s), or an extract. If an extract, the extract must be unambiguously defined. The inclusion of documentation to define the process by which the raw IPD is excluded from analysis was identified as desirable. This will enable the data requester to identify whether the research purposes would be best achieved by access to the raw IPD or the selected/curated IPD. This will also aid in informing the data quality statement for the project of the requesting researcher.

While participants were supportive of the use of IPD data standards (e.g. controlled vocabularies, common data models, etc) they acknowledged that complete standardisation of IPD was unlikely to be feasible or appropriate given the wide range of information types and test instruments used in clinical trials research. This is especially true for historical datasets, however there is opportunity to increase standardisation for new trials - especially by incorporating standards early in the research data lifecycle (e.g. built into electronic data capture systems) rather than post hoc to the trial at the stage that data is to be shared.

Most ethics approvals stipulate that to safeguard participants' privacy and confidentiality their IPD must be de-identified. This typically requires that identifying variables about participants, clinicians, and trial sites must be removed or coded, and careful attention must be paid to combinations of variables (e.g. age in years, sex, geographic locator, and population subgroup such as a rare disease phenotype) that may triangulate the identity of an individual participant. If such variables are critical to data re-use, then the IPD may need to be accessed in a controlled environment, or remote data processing may be required, to minimise the risk of re-identification.

Data quality requirements

Secondary users of data spoke to the value of accessing information about study quality, either from a data quality statement checklist, or from indirect measures of data quality included in the study protocol or other study documentation. The latter could include the use of data standards, a data management plan, a data monitoring plan (i.e. quality control measures, data validation steps) and a statistical analysis plan (including data exclusion criteria and validation process). As noted above, secondary data users also desire a measure by which data completeness can be assessed prior to requesting access to IPD.

It was routinely mentioned that HeSANDA could facilitate the improvement of data quality by providing multiple templates or guidelines for the design of clinical trials. This would need to be done in conjunction with both the SPIRIT and CONSORT guidelines, as well as conforming to the ethics bodies requirements for such documentation.

Data standards

Breakout session discussion identified that a number of data standards are currently in use. One or more participants reported using the following data standards to conduct their study and/or facilitate data re-use:

- CDISC[13] Foundational Standards (clinical research data)
- CDISC Exchange Standards (data sharing)
- SNOMED-CT[14] (electronic health records, including diagnoses, symptoms, procedures, pharmaceuticals, devices etc)
- DICOM[15] (images)
- MedDRA[16] (regulatory information)
- ICD[17] (diseases)
- WHO[18] (drug names)
- CTCAE[19] (adverse events)
- US FDA[20] (device names)
- Common data models ('CDM' - data schemas)

Participants did not express a desire that specific data standards should be required for the inclusion of study data in HeSANDA. However, they did suggest that if data is available in a standardised format then HeSANDA could provide additional value by implementing a data standards mapping to facilitate the transformation of data into differing standard formats.

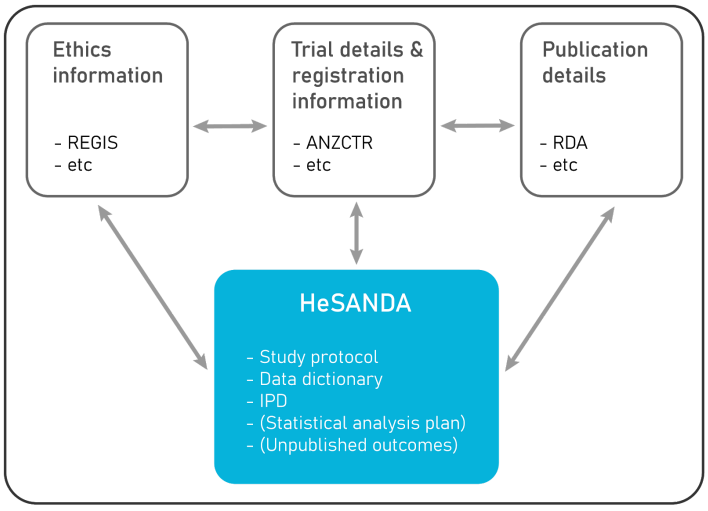
Draft HeSANDA information scope

Based on feedback received during the first two consultation workshops, a simplified schematic was drafted to represent the information needs of the research community and how HeSANDA could complement and integrate into existing research tracking systems. This draft 'information scope' (see Figure A.5 below) was presented in the third workshop to confirm ARDC's understanding of initial feedback and to refine requirements. Participants supported the draft information scope. Additional comments raised were to:

- Include metadata sharing of other key IPD sources such as images, omics, devices, etc;
- Coordinate seamlessly with other platforms to facilitate international collaboration;
- Consider that better education and training are required by all those involved in setting up clinical trials including the collection, analysis and storage of data;
- Enable control and access of IPD by the researchers who collect the original trial data; and
- Be future proof so that changes in the conduct of clinical trials (including dynamic consent), data technologies and data management practice can be integrated.

[13] Clinical Data Interchange Standards Consortium (https://www.cdisc.org/)
[14] Systematized Nomenclature of Medicine Clinical Terms (https://www.snomed.org/)
[15] Digital Imaging and Communications in Medicine (https://www.dicomstandard.org/)
[16] Medical Dictionary for Regulatory Activities (https://www.meddra.org/)
[17] International Statistical Classification of Diseases and Related Health Problems (https://www.who.int/standards/classifications/classification-of-diseases)
[18] World Health Organisation (https://www.who.int/teams/health-product-and-policy-standards/inn)
[19] Common Terminology Criteria for Adverse Events (https://ctep.cancer.gov/protocoldevelopment/electronic_applications/ctc.htm)
[20] U.S Food & Drug Administration (https://www.fda.gov/home)

Figure A.5. Draft HeSANDA information scope



Other requirements that were regularly identified by participants during the consultation series included:

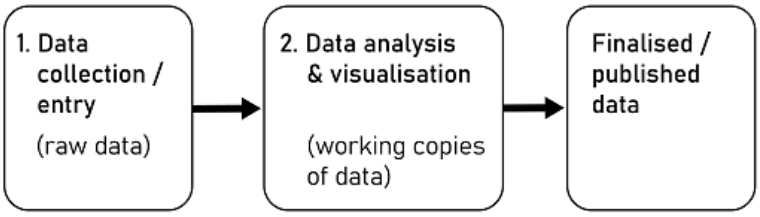
- The ability to query multiple studies using standard fields that covered the PICO framework (i.e. patient/problem/population, intervention, comparison/control, outcome)
- Guidance about preferred data standards by discipline
- Resources such as mapping between data standards
- Standardised templates for data documents
- Trusted ICT (i.e. information and communication technology) provider
- Proof of applicable accreditation
- Accessible and trusted data governance framework

Current and preferred data practices

In order to assess the feasibility of meeting the information needs expressed above, participants were surveyed on their current data practices. Since some participants were unavailable or had insufficient insight into these topics, an invitation to complete the survey was distributed to ACTA coordinating centres who are responsible for the oversight of data management for the majority of Australian investigator-initiated clinical trials. The results of this survey are reported below.

Note: For the purposes of the survey, a simplified 3-stage (IPD) data lifecycle schematic was used as a frame of reference in order to assess current practices during the research process:

Figure A.6. Simplified 3-stage data lifecycle



Clinical trial registration on a publically accessible website

The majority of people who conduct clinical trials indicated that they registered the trial on the ANZCTR (91%; 21 out of 23 respondents). Following the ANZCTR was ClinicalTrials.gov[21] (48%; 11/23), the World Health Organization's International Clinical Trials Registry Platform ('WHO ICTRP' - 17%; 4/23) and ISRCTN [22] (UK - 4%; 1/23). It is unclear why the WHO ICTRP was mentioned by respondents as it is a portal that collates trial registration details from primary registries and does not register studies itself. Some respondents indicated that they registered their trial on two or more registries (70%; 16/23). In relation to preferred or future practice, respondents indicated that the ANZCTR would be the main registry for trial registration.

Development and record of key clinical research documents

All of the people who are involved in conducting clinical trials indicated that they developed a study protocol (Table A.4). In addition, the majority of these trialists documented their terms or agreements for sharing IPD (72%; 13/18), data dictionaries (60%; 12/20) and statistical analysis plans (55%; 11/19). Recording unpublished study data was generally viewed as a low priority (17.6%; 3/17).

Table A.4. Data documentation by clinical trialists

Data recording or documentation?	Study protocol	Data dictionary with IPD	Statistical analysis plan	Unpublished study outcomes	Terms/agreements for sharing IPD
Yes	20 (100%)	12 (60%)	11 (58%)	3 (18%)	13 (72%)
Sometimes	0 (0%)	5 (25%)	7 (37%)	8 (47%)	0 (0%)
No	0 (0%)	3 (15%)	1 (5%)	6 (35%)	5 (28%)
Total	20	20	19	17	18

Use of templates or standard formats for key clinical research documents

It was common for people who conduct clinical trials to use a template or standard format for developing a study protocol (80%; 16/20; Table A.5). Workshop feedback indicated that SPIRIT and Transcelerate[23] generally guide the writing of a study protocol, so too do templates and standard operating procedures based on the requirements of the trial sponsor and coordinating centres.

Developing a data dictionary to accompany the IPD or a statistical analysis plan using a template or standard format did not appear to be part of routine practice (data dictionary: 47%, 9/19; statistical analysis plan: 37%, 7/19). Feedback highlighted that data dictionaries vary substantially from trial to trial and some standardisation for future trials has become a focus for certain organisations. Furthermore, there are significant differences in the skills and training across trial data managers and this can have an impact on the quality of data dictionaries, data structure and its storage. Data quality assurance was viewed as critical throughout the data life cycle as it helps to ensure reliable downstream analysis and meta-analysis.

[21] https://clinicaltrials.gov/
[22] International Standard Randomised Controlled Trial Number (https://www.isrctn.com/)
[23] https://www.transceleratebiopharmainc.com/

A recurring comment at the workshop was that minimum standards for data sharing and transfer agreements as well as templates and tools for data management practice (including better data transfer methods) need to be developed.

Table A.5. Use of templates or standard formats by clinical trialists

Template or standard format used?	Study protocol	Data dictionary with IPD	Statistical analysis plan
Yes	16 (80%)	9 (47%)	7 (37%)
Sometimes	1 (5%)	2 (11%)	6 (32%)
No	3 (15%)	8 (42%)	6 (32%)
Total	20	19	19

IPD collection and entry[24]

The most frequently used technologies to collect and enter IPD by clinical trialists were REDcap (25%), 'other' technologies (18%) and Microsoft Excel (18%; Table A.6). 'Other' technology was not defined in the survey or by survey respondents. The average number of technologies in use were three per respondent; there were 24 respondents for this question. In relation to preferred technologies, respondents indicated that REDCap was their top choice.

Workshop discussions identified that some clinical trial teams use bespoke data collection and management tools created in-house or are in the process of adapting current tools. These electronic data capture systems may need to adhere to other standards such as those outlined by the US FDA. Further to this, some feedback indicated that for smaller studies REDCap would suffice.

Table A.6. Technologies used for IPD collection/entry

Technology	Currently use	Prefer/plan to use
REDCap	19 (25%)	10 (43%)
Other*	14 (18%)	4 (17%)
Microsoft Excel	14 (18%)	1 (4%)
Microsoft Access	8 (11%)	1 (4%)
Qualtrics	7 (9%)	2 (9%)
Survey Monkey	6 (8%)	2 (9%)
API requests	3 (4%)	1 (4%)
SQL queries	3 (4%)	2 (9%)
NoSQL queries	2 (3%)	0 (0%)
Total	76	23

*Not specified in the survey or by respondents

[24]For the following sections on software and file formats, the reader is directed to the following online resources for further information on the terms used: Lists of software & List of file formats

IPD analysis or visualisation

The top five technologies for IPD analysis or visualisation were SAS, R, STATA, Other and SPSS irrespective of whether the respondent was collecting IPD for a trial (Figure A.7) or re-using IPD for research purposes (Figure A.8). 'Other' was undefined in this survey. The average number of technologies in use was, on average, 3.5 technologies for data collectors (22 respondents) and 3.3 technologies for people who re-use existing data (15 respondents). In relation to preferred technologies, respondents indicated that they would use a variety of technologies and these were generally the same as those that are currently being used.

Of note, workshop feedback and data from the survey suggests that R, SAS and Python are gaining popularity. For researchers who collect and analyse medical image data, important technologies include Python and MATLAB.

Figure A.7. Technologies used for IPD analysis &/or visualisation for a trial

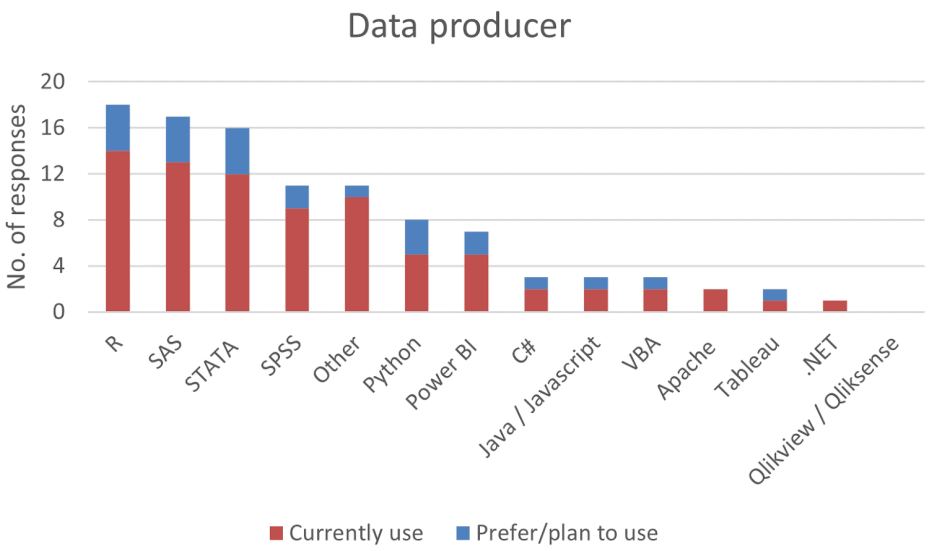
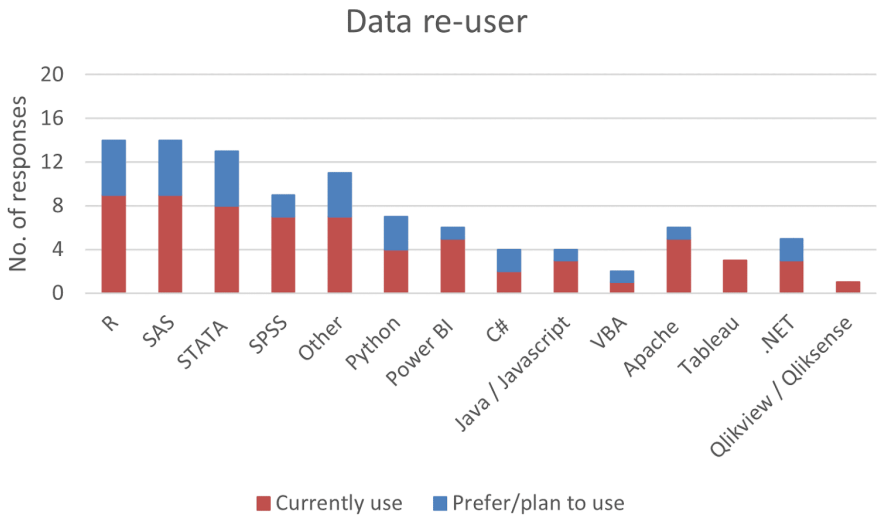


Figure A.8. Technologies used for IPD analysis &/or visualisation for data re-use



IPD formats

The most common data formats currently used for raw data were CSV/PSV, SPSS, XLS, SAS and non-data formats (e.g. DOCX, PDF, MPG, JPG, etc) for people who run and collect participant data for a trial (Figure A.9). For working copies of data, people frequently used CSV/PSV, XLS, SPSS, STATA, SAS and XML; while for the finalisation and publication of data, the format was typically non-data format (e.g. PDF), STATA, CSV/PSV and SPSS. Each respondent indicated that, on average, six data formats would be used when working with IPD. Based on workshop discussions, researchers who work with medical imaging and treatment data use DICOM format and processing tools such as 3D Slicer, ImageJ, ITK-SNAP and MATLAB. Respondents who re-use IPD were capable of using non-data formats, Access, SAS, XLS, SPSS and CSV-PSV, while their preference would be to use CSV-PSV, STATA or SAS (Figure A.10). Each respondent indicated that, on average, six data formats could be used when re-analysing data.

Figure A.9. IPD formats used when collecting clinical trial data

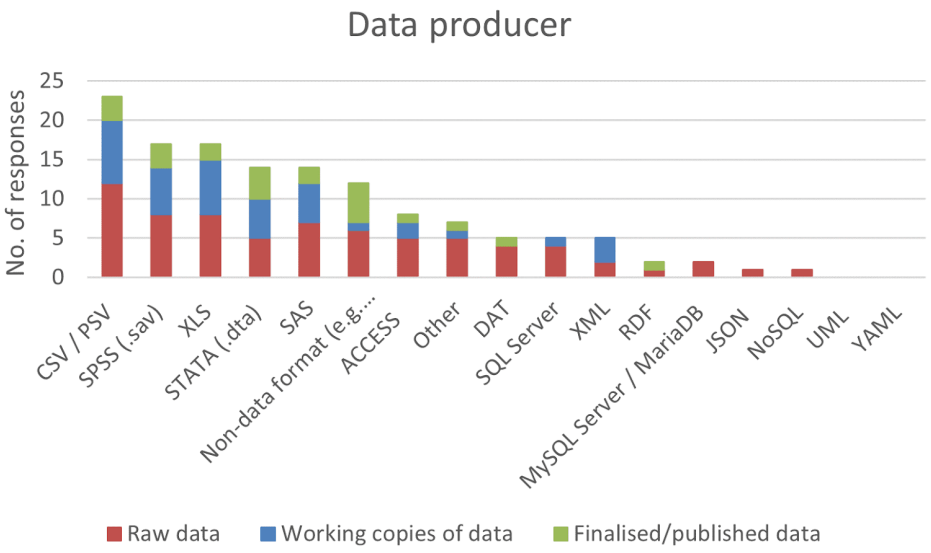
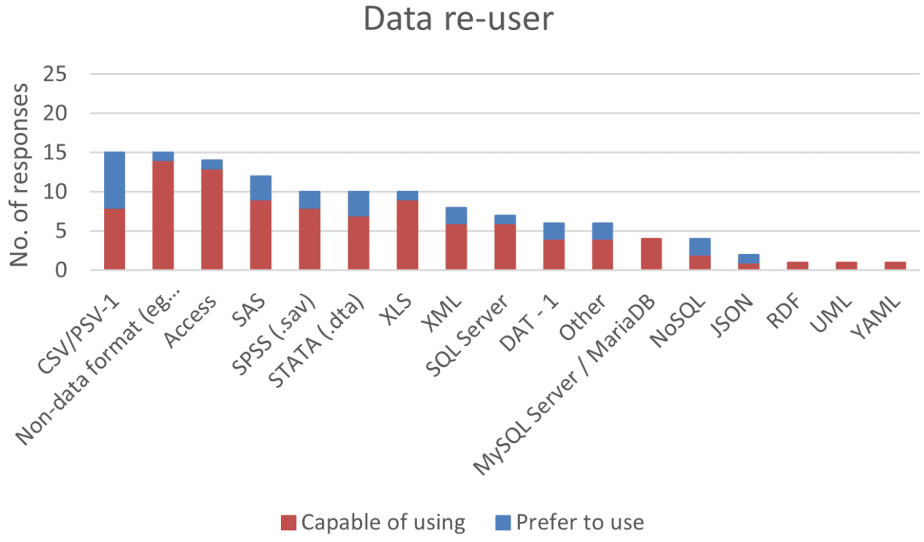


Figure A.10. IPD format when re-using trial data



Methods or terminologies to describe or code data

NB. This survey question probed for details around current data standards. The response options reported represent different types of standards (e.g. controlled vocabularies and classification systems, documentation standards, information exchange protocols, etc) rather than single type.

For people who collect original trial IPD, the most commonly used method to describe data in current practice was the data dictionary (Table A.7). This was followed by ICD (currently ICD-10) and 'Other', however the latter remained undefined. On average, each respondent used three or more methods or terminologies to describe IPD based on 24 respondents. In relation to preferences, respondents highlighted the potential use of data dictionaries, CDISC (CDASH/SDTM etc) and common data model.

For people who re-use trial IPD, the responses were very similar with the data dictionary being the preferred method followed by ICD-10, common data model and metadata registry (Table A.7). On average, respondents selected three methods or terminologies as part of their preferred methods and this was based on 14 respondents. The responses were generally consistent across the two types of respondents.

Table A.7. Methods or terminologies for IPD

Method	Data producer (*n* = 24)		Data re-user (*n* = 14)
	Currently use	Prefer/plan to use	Preferred methods
Data dictionary	**20 (83%)**	**6 (25%)**	**13 (93%)**
ICD-10	6 (25%)	2 (8%)	6 (43%)
CDISC (CDASH/SDTM etc)	3 (13%)	4 (17%)	2 (14%)
Common data model	3 (13%)	4 (17%)	5 (36%)
Other*	6 (25%)	1 (4%)	2 (14%)
Data scoping/authorisation documentation	4 (17%)	1 (4%)	3 (21%)
Metadata registry	3 (13%)	1 (4%)	5 (36%)
SNOWMED-CT	2 (8%)	1 (4%)	2 (14%)
MESH	2 (8%)	1 (4%)	4 (29%)
FHIR	0 (0%)	1 (4%)	0 (0%)
HPO	0 (0%)	0 (0%)	0 (0%)

*Not specified in survey or by respondents

Data storage, archiving and publication

IPD were typically stored in institutional services/shared drives and formal repositories or archives. Key data documentation including study protocols, data dictionaries, statistical analysis plans and terms or agreements for sharing IPD were located in institutional servers, followed by formal repositories or archives, peer-review journals and clinical trial registries. Current practice indicated that unpublished study data are

generally kept in institutional servers/shared drives and somewhat in format repositories or archives (Table A.8).

In relation to preferred or future practice, respondents indicated that they would prefer to use a formal repository for storing IPD and all key documents. This was a marked difference from current practice (currently: 19% vs future preferred practice: 81%). Feedback points from the workshop included that current data storage practice is very siloed and institutional databases are commonly used rather than other locations for data storage. The value of safe havens or 'walled gardens' for data storage and re-use was highlighted during workshop discussions.

Table A.8. Current and preferred practice in storing data and key documents

Storage, archiving or publication site	Study protocol	IPD	Data dictionary	Statistical analysis plan	Unpublished study data	Terms for sharing IPD	Total
Institutional servers/shared drives							
Current	19 (17%)	20 (18%)	20 (18%)	18 (16%)	16 (15%)	17 (15%)	110
Preferred	3 (9%)	4 (11%)	7 (20%)	7 (20%)	8 (23%)	6 (17%)	35
Peer-reviewed journal							
Current	14 (52%)	N/A	3 (11%)	9 (33%)	N/A	1 (4%)	27
Preferred	7 (35%)	N/A	5 (25%)	6 (30%)	N/A	2 (10%)	20
Clinical trial registries							
Current	8 (47%)	1 (6%)	1 (6%)	4 (24%)	1 (6%)	2 (12%)	17
Preferred	6 (23%)	5 (19%)	3 (12%)	3 (12%)	5 (19%)	4 (15%)	26
Formal repository or archive							
Current	5 (45%)	2 (18%)	1 (9%)	2 (18%)	1 (9%)	0 (0%)	11
Preferred	7 (15%)	10 (21%)	11 (23%)	7 (15%)	7 (15%)	5 (11%)	47
Other							
Current	2 (17%)	4 (33%)	1 (8%)	3 (25%)	2 (17%)	0 (0%)	12
Preferred	1 (17%)	2 (33%)	1 (17%)	1 (17%)	1 (17%)	0 (0%)	6

Methods to access clinical trial data and documents for secondary use

The most frequently used method to access clinical trial data and key documents was by directly contacting the clinical trialist (37%; 86/235) followed by searching a clinical trial registry (18%; 42/235), or accessing a research repository (16%; 38/235) or peer-reviewed journal (15%; 35/235; Table A.9). Clinical trial registries, research repositories and peer-reviewed journals were predominately used to source general trial details, study protocols and statistical analysis plans. Notably, 17 people responded to this section of the survey

relating to current practice for people who re-use data. Frustrations were noted by researchers who are trying to access data. This was because:

- there is no single source to access data (aside from clinical trial registries);
- some organisations may be uncooperative and wish to maintain data control;
- study protocols are not easily accessible and ethics committees are unresponsive or slow, and not focussed on data re-use; and
- the HREC approval and participant consent forms from the trial may not have the appropriate approvals in place to support the secondary use of data.

Of note, it was suggested that good practice in clinical trials needs to be biased towards data sharing and that data from investigator-initiated trials would need to be available to all.

Table A.9. Current practice to find and access data and documents for data re-use

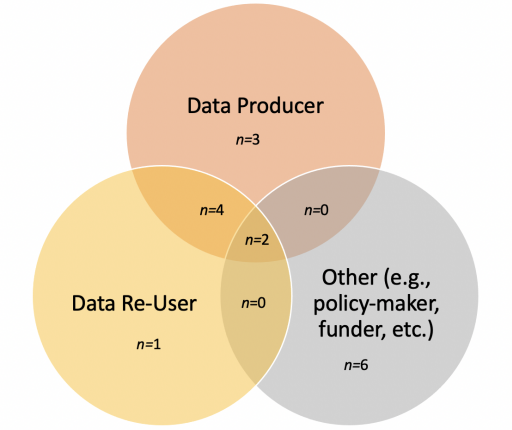
Methods used to access data[1]	General trial details	Study protocol	IPD	Data dictionary	Statistical analysis plan	Unpublished trial	Data sharing terms or agreements	Total
Contact with trialist	**15 (17%)**	**14 (16%)**	**12 (14%)**	**11 (13%)**	**11 (13%)**	**14 (16%)**	**9 (10%)**	**86**
Clinical trial registry	**15 (36%)**	**11 (26%)**	2 (5%)	3 (7%)	5 (12%)	5 (12%)	1 (2%)	42
Research repository	**8 (21%)**	**8 (21%)**	5 (13%)	4 (11%)	5 (13%)	5 (13%)	3 (8%)	38
Peer-reviewed journal	**12 (34%)**	**9 (26%)**	4 (11%)	1 (3%)	8 (23%)	0 (0%)	1 (3%)	35
Not tried	0 (0%)	1 (5%)	4 (18%)	4 (18%)	4 (18%)	2 (9%)	7 (32%)	22
Other	2 (17%)	2 (17%)	2 (17%)	1 (8%)	1 (8%)	1 (8%)	3 (25%)	12

Barriers & incentives for data sharing

The focus of the final consultation workshop was to identify the incentives and barriers for data sharing, as well as the potential enablers that might address them. The workshop sought to elicit participants' views on the issues that were most likely to affect their decisions to engage both in data sharing (data producers) and the secondary use of data (data users). Feedback was initially obtained via live polling during the main presentation section of the workshop - the results of these polls are presented immediately below. Participation in the initial polling was modest, but was utilised to initiate consideration of the range of issues which was then explored in depth with greater participant engagement during the subsequent workshop breakout session. It should be noted that many of the issues had already been raised by participants during previous workshops. The summation and consensus of opinions expressed in the breakout session and across the entire consultation series is reported in the Summary of Feedback section below.

Participants (*n*=20) were first asked to identify their role(s) across three options: “data producer”, “data re-user” or “other (e.g., policy-maker, funder, etc.)”. These categories were not mutually exclusive and participants were encouraged to nominate all applicable categories. Sixteen of the 20 participants responded to this question. Figure A.11 shows that nine participants identified as data producers: of these, four people indicated that they were also secondary users of data, and two indicated both data re-use and “Other” roles. Six participants identified exclusively as “Other” and one identified exclusively as a data re-user. Thus, despite the modest number of participants, all three roles were represented, and with some substantial overlap.

Figure A.11. Participants’ Role(s) (*n = 16)*



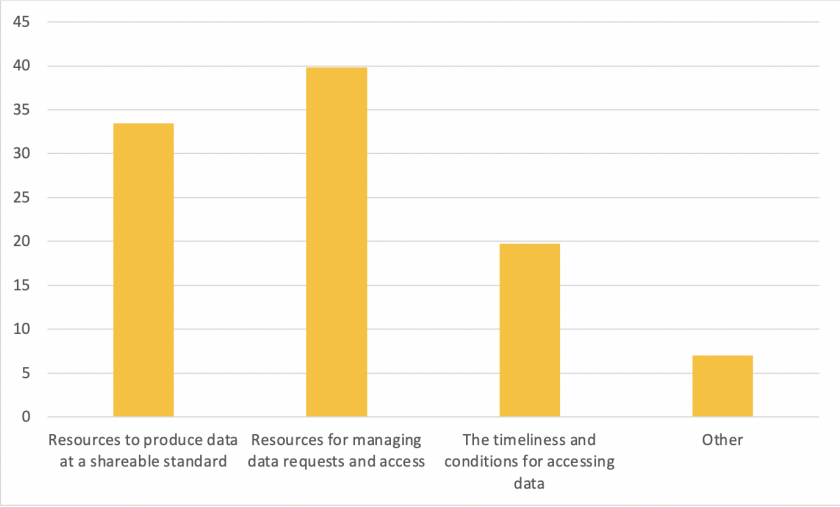
Participants were then asked to respond to four questions aimed to identify their concerns around the feasibility of engaging in data sharing. For each question, participants were given 100 points to allocate across the predefined response options. They could allocate as many or as few points to each item as they considered was warranted by its importance. Participants could also choose to allocate less than 100 points in total (or no points at all) if they considered items to be of little or no importance.

The predefined response options were based on feedback received during the preceding workshops and a response option of “Other” was included for each question to record concerns that were not adequately described by the predefined responses. The breakout session that followed enabled in-depth discussion of participants’ concerns and the capture further information about responses under the category “Other”. As mentioned, these discussions are reported in the Summary of Feedback section which follows the polling results.

Costs & feasibility

The results in Figure A.12 show that, of the cost and feasibility issues on which they were asked to cast votes, participants were particularly concerned about the resource requirements that may be involved to manage data requests and access and to produce data at a shareable standard. Timeliness and the conditions for gaining access to data were also considered important and approximately seven per cent of votes were cast for other feasibility and cost issues.

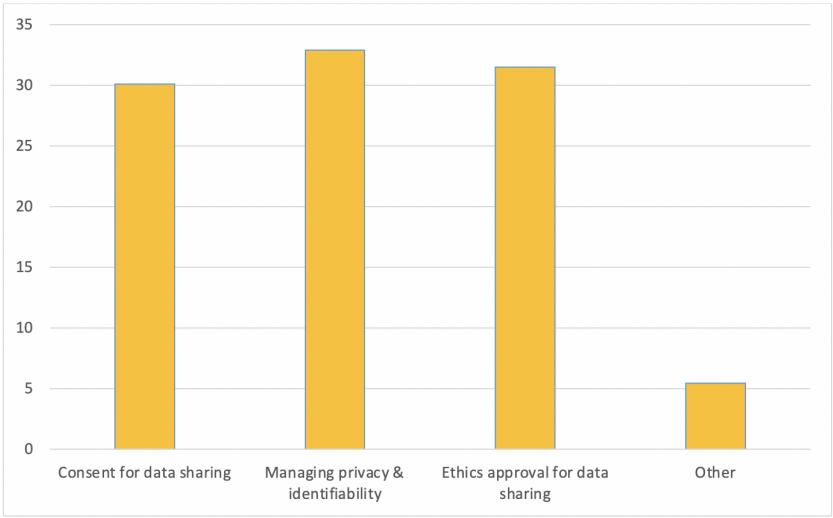
Figure A.12. How Much Do These Cost and Feasibility Issues Impact You? *(n = 18)*



Ethics

Figure A.13 reports the responses to a question about the privacy- and ethics-related concerns that were considered most important to participants. Notably, each of the three nominated categories concerning consent, the management of privacy and identifiability, and gaining ethical approval for data sharing attracted 30 per cent, or more, of the total votes. “Other” issues received approximately 5.5 per cent of the votes cast.

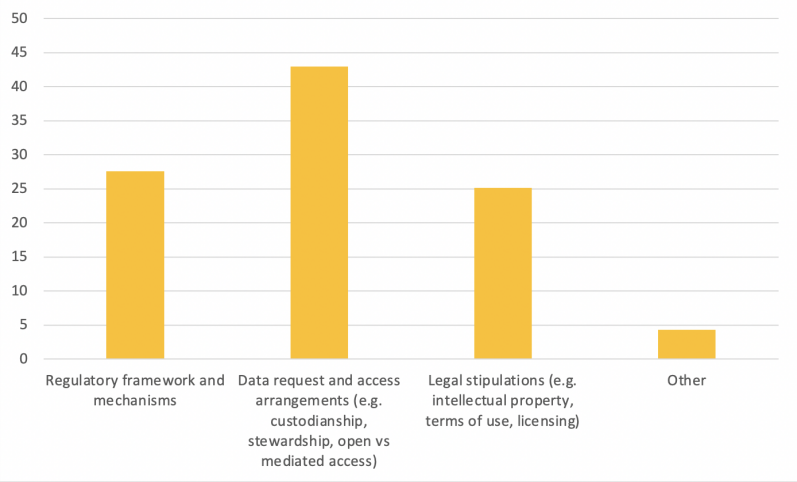
Figure A.13. How Much Do These Ethics Issues Impact You? *(n = 19)*



Data governance

Figure A.14 presents participants' voting responses on a question about the relative importance of governance-related considerations. Approximately 43 per cent of all votes were cast on data request and access arrangements; while the regulatory framework and mechanisms and legal matters attracted approximately 28 and 25 percent of votes, respectively. Approximately four per cent of votes were cast on the "Other" category.

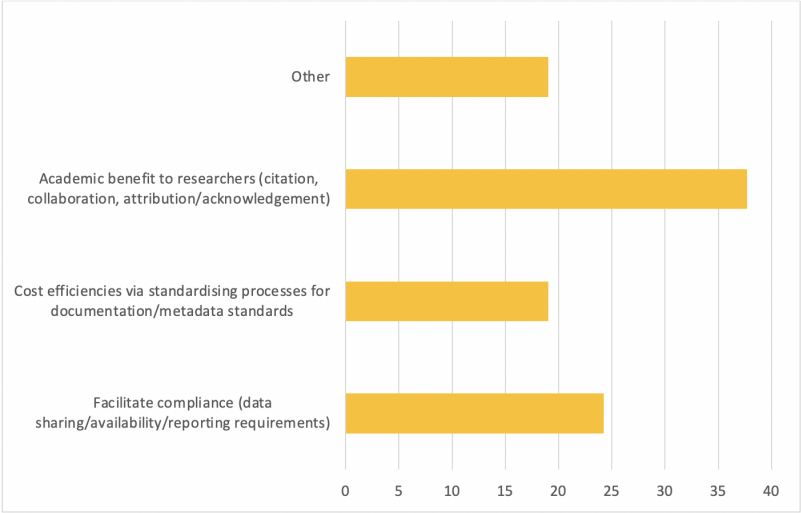
Figure A.14. How Much Do These Governance Issues Impact You? *(n = 19)*



Incentives

Figure A.15 shows that participants considered the strongest incentives to share data via the HeSANDA initiative were concerned with various sources of academic benefit to researchers, while efficiency and compliance issues also attracted substantial proportions of the total vote. It is noteworthy that the "Other" category in this case also attracted a substantial proportion (approximately 19 per cent) of the votes. It became evident through comments during the polling session and in the breakout sessions that the latter result owes specifically to the weight that a number of participants placed on the public (or social) good that would be produced by effective and efficient data sharing and data re-use.

Figure A.15. What Are the Biggest Incentives to Share Your Data Via HeSANDA? *(n = 19)*



Summary of feedback

In this section, the feedback received on the specific topics above has been harmonised and contextualised within the framework of the four main consultation themes.

Theme A: Research purpose & value proposition

The workshop participants were predominantly researchers and clinical trialists, however the representation was diverse. The workshop and survey format allowed for the opinions of a wide range of stakeholders to be gained. Many participants identified as both potential data users of, and data contributors to, the initiative.

Consensus regarding the research purposes of data sharing highlighted need in three principal areas:

1. Secondary research projects and analyses
2. Meta-analysis and systematic review
3. Replication, reproducibility, or peer review

The consensus view was that of a strong desire to move towards the standardisation of research outputs through all aspects of the trial process (from data collection through to study reporting) which could facilitate the efficient sharing and secondary use of clinical trials data.

While many participants had limited experience with current data sharing platforms, the feedback received indicated that they do not meet the range of data content and quality requirements necessary to efficiently achieve the identified research purposes. Participants believed that to maximise utility, data sharing infrastructure must have extensive search function capability and provide a single, safe and reliable source of primary information about clinical trials.

Participants highlighted the value of coordinating the delivery of tools, services and supporting resources (e.g. standard ethics templates) that support data sharing in order to provide efficiency gains in the three principal areas described above.

Theme B: Data content & quality requirements

The workshop participants expressed clear consensus to prioritise access to IPD for secondary research use. Participants also considered the study protocol, data dictionary, proof of ethical approval and consent to data re-use as essential for effective secondary use of IPD. Participants expressed a desire for indicators of data quality/completeness but did not identify a preferred indicator.

Other summary information about clinical trials was considered important but participants noted that this information is already collected during trial registration and ethics approval processes. They indicated a desire to utilise these existing sources of trials information to avoid duplication and support the coordination of research tracking systems. Consequently, an Information Scope for HeSANDA was drafted to reflect the reported information needs, the existings sources of this information, and the potential integration and value HeSANDA could bring to this landscape.

Participants indicated that there was general alignment of the draft HeSANDA information scope to their research needs and existing data practices, and suggested that the infrastructure should support as many data file formats and types as possible (e.g. big data formats such as medical imaging and genomics) and ideally should be interoperable with existing and emerging international research data systems and infrastructure.

Theme C: Existing data standards & practices

At present, clinical trials do not consistently adopt data standards and those that do use a variety of standards and data practices. The consensus view was that the use of a defined set of standards and practices should be encouraged in new studies but the feasibility of implementing these in the initial rollout of HeSANDA (e.g. for the sharing of existing trials data) will require further assessment.

IPD standards are also variable due to the myriad data types and formats collected during trials. Although study protocols, data sharing agreements, data dictionaries and statistical analysis plans were typically produced when conducting a clinical trial, the use of templates or standard formats was low. Notably, templates were used when developing a study protocol but not routinely for other key documents. It was evident that some agreed approach for recording data dictionaries, statistical analysis plans, and data sharing agreements would be beneficial to the research community.

Individuals who are involved in running clinical trials or re-using clinical trial data consistently highlighted REDCap as a data collection tool, but specific organisations use bespoke data collection systems to comply with data standards set by trial sponsors, trial coordinating centres, or the US FDA. Regardless of specific systems or software tools, data dictionaries were regularly developed to describe the IPD variables.

There is a wide variation in the technologies used to analyse IPD by people who conduct trials and those that re-use the data but R, SAS and STATA are prominent technologies amongst researchers.

There was a notable preference to store IPD and key documents (protocol, data dictionary, statistical analysis plan, unpublished study data, and terms for sharing IPD) in a formal repository or archive.

A recommended practice was better educational materials and training in data management. In addition, both clinical trialists and secondary users of data agreed there would be benefits in standardising templates and other resources used in study design, data collection, and trial reporting.

Theme D: Barriers, systems, & enablers

Of the various feasibility matters discussed throughout the consultations, ethical and privacy issues were unambiguously the primary concern for many participants, and these were inexorably linked to the other feasibility concerns. Specifically, the consensus view was that ethical and privacy concerns drive many of the governance-, cost- and other feasibility-related concerns that participants had. This related set of concerns were considered the most important potential barriers to be addressed for effective and efficient data sharing and re-use.

Data producers also emphasised the currently-high costs of fielding and fulfilling data requests. They identified this as an important potential disincentive to high levels of participation in data sharing. Similarly, secondary users reported substantial transaction and (in some cases) monetary costs of gaining access to data for secondary use. Addressing these resource costs of managing and sharing data was considered pivotal in enabling productive data sharing practices.

Attendees agreed on some key areas for development that would help address these barriers:

- Attendees expressed a strong desire for guidance and/or standards to address ethics requirements to enable data sharing. For example, workshop attendees recommended the development of a common standard for participant consent that, when incorporated in data collection protocols, could specify clear conditions and provide data custodians with confidence that the data may subsequently be shared. Attendees agreed that it would be important to provide clarity on (i) the types of data that may be shared and (ii) the conditions under which this could occur. The establishment of such

guidelines that could be endorsed by major stakeholders (e.g., the NHMRC) and promulgated nationally (particularly among Ethics Committees) to streamline data-sharing was seen as an ideal goal.

- Attendees also desired guidance and resources on data standards and practices that can assist data custodians to organise and store their data in a manner that, prospectively, will lower the cost of data sharing. One common suggestion was the provision of a best practice guide and templates that assist researchers to store their data and metadata in a way that will aid data sharing without the need for extensive post hoc manipulation of large datasets.
- The provision of data to secondary users via trusted research environments ('TRE', aka 'secure research environments', 'remote access data laboratories', 'virtual workbenches') was also recommended as a way to provide confidence to researchers that data access would meet the privacy concerns of research participants and ethics committees. The integration of trial metadata catalogues, data repositories, a standardised data access framework, and TREs was suggested as an ideal ecosystem for enabling efficient data sharing with robust regulatory compliance.

Discussion

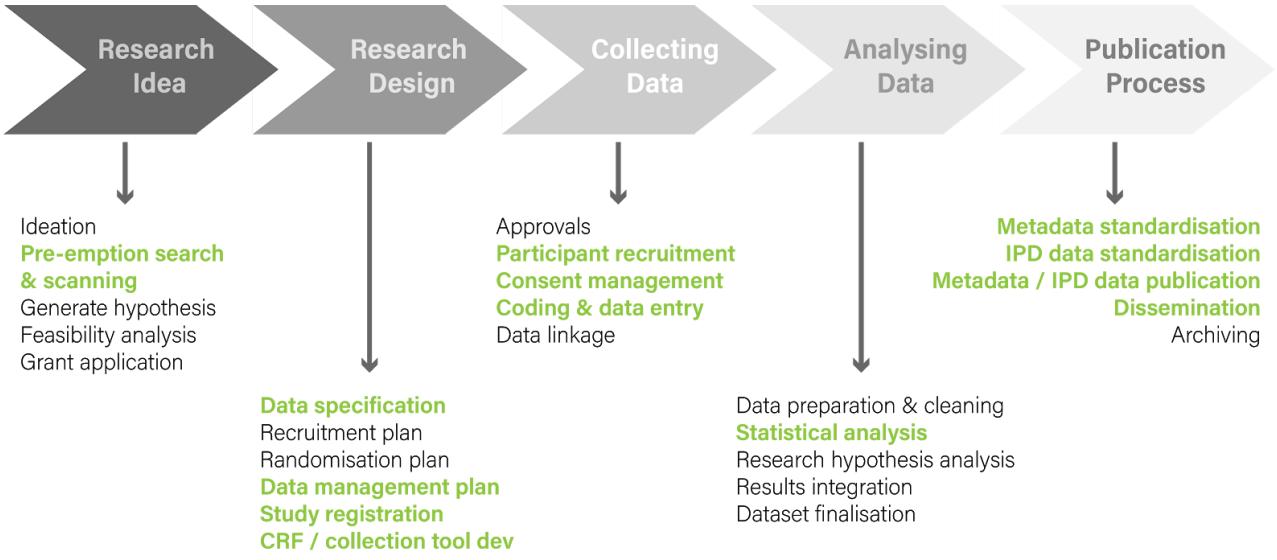
The consultations established that:

1. Australian researchers represented in the consultations are enthusiastic about the HeSANDA initiative: both data producers (including the clinical trialists and participants) and secondary data users expressed support for HeSANDA's vision. In particular, consultation participants emphasised HeSANDA's value as a vehicle for sharing and accessing high-quality IPD and associated metadata. Researchers reported strong incentives to engage with HeSANDA that pertained not only to enhancing their own professional activity, but with **maximising the public good** that could be achieved by developing a more efficient approach to data management and sharing.
2. There is **currently a lack of standardisation** in the management of clinical trial data both with respect to the data systems and practices that different research groups use, as well as the implementation of data governance by different custodians and jurisdictions. Indeed there is uncertainty around the ethical and consent requirements regarding secondary data use that presents a barrier to researchers wishing to engage in data sharing. HeSANDA will need to take an active role in addressing not only issues of data management practices and data systems and data standards but also data governance in order to support successful data sharing.
3. Despite community enthusiasm for data sharing and secondary use, **the high costs of standardising and sharing data was reported as a barrier.** Similarly, for secondary data users, the costs of finding and securing the use of existing data are currently also very high. These include the resource costs associated with searching for suitable data when the publicly-available metadata are often insufficient to determine whether a specific dataset is fit-for-purpose for the secondary user.
4. **HeSANDA can play an important enabling role in supporting Australian health research efficiency, impact, and translation by addressing these issues at the national level** and promulgating standards that will make data sharing and secondary use more efficient. This includes, but is not limited to, adopting a leading role in standardising both data governance frameworks (including participant consent) and also the conventions and mechanisms that are adopted nationally for data sharing and secondary use.

Responses elicited throughout the consultation identified generalisable stages and requirements of the research journey for a clinical trial. The research journey and how this intersects with data and secondary use is shown in Figure A.16 below. Highlighted in green are requirements of the research journey that

intersect with the high-level goals of HeSANDA as determined through the consultation process and the consolidation of the workshop themes:

Figure A.16. The Research Journey and intersections with data and data use



Principles

Based on the feedback obtained via the consultation process, a list of principles can be derived for HeSANDA which reflect the needs and perspective of the clinical trials research community who participated in the initial data development workshops. These principles can provide direction that will assist HeSANDA to achieve its vision and mission to promote data sharing and secondary use in Australia. The workshop feedback informing each principle is listed in the 'source' column.

Principle	Statement, rationale, further information	Source
Purpose		
[1] The capabilities delivered by HeSANDA must be informed by the core value proposition	HeSANDA will enable the national infrastructure required to support the sharing and secondary use of health research data – improving research efficiency, reducing cost, and increasing research impact.	Theme A
[2] The core research purpose of HeSANDA is to support research with a translational focus	HeSANDA will support access to the information and outputs from clinical trials necessary for: ● meta-analysis and systematic review ● replication, reproducibility, &/or peer review ● secondary research projects and analyses to facilitate the translation of research into clinical guideline and policy development, health technology assessment, and the development of new research.	Theme A

[3] HeSANDA will facilitate the sharing of a range of clinical trial information	To meet the needs of data producers and secondary users, HeSANDA will support the sharing of a variety of different types of information associated with clinical trials, with an emphasis on individual participant-level data, study protocol metadata, and cohort summary data.	Theme A, C
[4] HeSANDA will maximise the discoverability of the clinical trial information	The information available through HeSANDA must be organised in a way that supports efficient search and discovery of clinical trial information (e.g. using the PICO framework).	Theme A, B
[5] HeSANDA will improve the efficiency and reliability of access to clinical trial data for secondary research	Trial information is currently siloed, predominantly stored on institutional servers, and often accessible to secondary researchers only via direct contact with individual trialists. However, there is a clear community enthusiasm for making this information accessible via more standardised and potentially centralised mechanisms to achieve optimal research efficiencies.	Theme C
[6] HeSANDA will minimise the barriers to data sharing by clinical trialists	In order to reduce resource costs and facilitate development of the data asset, HeSANDA must align data sharing with existing research practices.	Theme D
Data content & quality		
[7] HeSANDA will promote minimum reporting & data sharing requirements for clinical trials	The implementation of minimum requirements maximises the utility of clinical trials data. Minimum data sharing requirements must include IPD and the study information that contextualises it (i.e. study protocol; data descriptions; data quality statements). Research and data descriptions to support Principle 4 must also be included. Enabling coherent data practices throughout the research journey can support these requirements.	Theme A, B, C, D
[8] HeSANDA will support the current variety of IPD data standards but will encourage pathways to the adoption of stakeholder-endorsed data standards	Currently, there is a wide variation in the data formats used to collect, enter and analyse new data. As such, HeSANDA will need to facilitate the sharing of different data formats for IPD and metadata. However, HeSANDA should support the adoption of standardised data platforms and data standards for storing data and recording metadata (e.g., data dictionaries).	Theme C, D
[9] Data quality statements will underpin the utility of HeSANDA's content	In order to provide confidence in the data asset, data quality should be represented for each data collection.	Theme A, B, C

Data governance		
[10] HeSANDA should promote common approaches to data sharing and re-use by clinical trials researchers	Researchers encounter resource and efficiency issues due to the lack of clear guidance on how to implement the data sharing policies of funders, publishers, and other stakeholders. The development of agreed protocols and procedures will improve the feasibility for data sharing to become standard research practice.	Theme A, C, D
[11] HeSANDA should promote common approaches to participant consent requirements for data sharing and re-use	Researchers agree on the fundamental importance of consent and community support for research practices such as data sharing. But as with the previous principle, they require guidance on how best to implement open science policies as they relate to participant consent. Developing a coordinated national approach to meet consent requirements will not only improve the feasibility of data sharing but, most importantly, address the concerns and mitigate risk around the sharing of sensitive data.	Theme A, D
[12] HeSANDA should promote best practice guidelines for the handling and sharing of sensitive data	To complement the principles of common approaches to policy interpretation and application (above), researchers will benefit from guidance on specific data handling issues such as data de-identification, security, etc.	Theme A, C, D
[13] HeSANDA should be considerate of the labour cost to clinical trialists to facilitate access to data	The above principles seek to improve the efficiency of data sharing (either directly or indirectly), thereby reducing costs and improving feasibility. However, these improvements cannot entirely remove the labour cost of data sharing that is not consistently supported at the funder or institutional levels at present. Recognition of these costs within data sharing policy and infrastructure is fundamental to supporting the research community.	Theme A, D
Stakeholder coordination		
[14] HeSANDA should align its activities with existing structures and initiatives that support the national harmonisation of clinical trial activities	For example, currently clinical trial researchers are required to enter common data regarding their trial in the human research ethics application (HREA) form, trial registration (e.g. via ANZCTR), and, where applicable, to the Therapeutic Goods Administration (TGA). To reduce administrative burden for researchers, HeSANDA will link to these and other existing structures to support better knowledge discovery and easier meta-analysis.	Theme A, B, C

[15] HeSANDA should attempt a nationally coordinated approach to address its data governance aspirations and principles	Issues of data sharing & governance impact multiple stakeholder groups, from research participants and researchers through to funders, institutions, and ethics committees. The research community desires cooperation and coordination between these groups to address their common interests.	Theme A, D
[16] HeSANDA should leverage existing investment in data sharing infrastructure where possible	Researchers are required or incentivised to utilise existing data management and sharing infrastructure provided by their organisations. HeSANDA should engage with research organisations in order to develop strategies to avoid unnecessary duplication of effort and to maximise existing infrastructure investments.	Theme C, D

Recommendations

The principles above can be combined with the stages of the research journey of relevance to HeSANDA to identify specific areas that could form the basis of effective infrastructure investment. The stages and potential areas for investment (as recommended by the editorial team) are detailed below:

Phase	Stage	Areas for potential investment
Research ideation	Pre-emption search and scanning	• Centralised research/data discovery tools
Research design	Data specification	• Determination, advocacy and advancement of the best data standards to underpin clinical trials • Tools to support the easier adoption of data standards at the trial design phase • Also, see 'Metadata standardisation' and 'IPD standardisation' below
	Data management planning	• Research community support and development of data management plan (DMP) templates and standards to facilitate data sharing • Enhancement of electronic DMP systems • Promote and facilitate the development of standard data sharing accords and agreements to maximise researchers' ability to undertake research • Promote the use of data management systems and environments that support data security best practice
	Study registration	• ANZCTR/TGA or other registry metadata enhancement and standardisation • Centralised research discovery tools or links
	CRF / collection tool development	• Support the development of trial data collection tools to enhance data standards conformance
Collecting data	Participant recruitment	• Development and availability of templates for consent that considers data re-use • Central guidance resources for researchers to support data sharing aspirations
	Consent management	• Consideration of guidelines, platforms, and/or tools that support better informed consent processes (for example, dynamic consent)

		• Establish key stakeholder agreement on a common approach to consent management, ideally with endorsement at a national level from NHMRC and/or HRECs
	Coding and data entry	• Promoting the availability of CRF / collection tools that utilise data standards appropriate to clinical trials
Analysing Data	Statistical analysis	• Promote the availability or sharing of standard data analytic code and procedures (for example, statistical code scripts or common data model analytic packages) • Promote the availability or sharing of standard analytical procedures, syntax, and informatics libraries • The use of TREs to maximise data security
Publication process	Metadata standardisation	• Support a community of practice to achieve consensus around metadata standards • ANZCTR / TGA or other registry metadata enhancement and standardisation
	IPD data standardisation	• Consult with IPD experts (nationally and internationally) to elucidate the minimum standards of data structure for IPD • Support communities of practice and community consensus around appropriate standards • Support training and uplift in applying standards • Support the enhancement of the ability to map datasets between different data collection formats and conversion of data to standards that can support data re-use (for example, CDISC, CDMs, etc) • Mechanisms for publication of data mappings • Tools for data conversion
	Metadata / IPD data publication	• Provision of clinical trials metadata and IPD publication capabilities in conjunction with university and research institute infrastructure (centralised and federated components)
	Dissemination	• Development of a community of practice and materials to develop and support concepts such as computable research study results
	Data sharing	• Promote and facilitate the development of standard data sharing accords and agreements to maximise researchers' ability to undertake research • Increasing the efficiency of data governance frameworks relating to data sharing and access • Establish stakeholder-endorsed data request and approval mechanisms • The use of TREs to provide controlled secure access to data

These recommendations can be distilled into a set of three key priorities for infrastructure development:

1. A set of **coordinated data services** that:

    - Facilitate access to IPD for secondary use
    - Facilitate access to study summary information, protocols, data dictionaries, data quality statements, and ethics information to enable research discovery and secondary use of data
    - Support common data and metadata standards
    - Supply standardised descriptions to central discovery services (not held elsewhere)
    - Provide access to data according to a common governance framework
    - Support centralised data request and access processes
    - Provide tools for researchers to efficiently meet the above requirements

2. A set of **federation services** that integrate the coordinated data services to enable:

    - Research and data discovery
    - A streamlined data request process
    - Efficient data access

3. A set of stakeholder-endorsed **coherent data practices** for:

    - Research data and metadata standards
    - Standardising compliance with ethics and participant consent requirements
    - Data governance
    - Data request and access processes
    - Tools to facilitate data standardisation and compliance

These services and practices should adhere to two key requirements:

- Data sharing should support the interests of data producers and secondary users, research participants and the general public, research institutions and organisations, funding agencies and policy makers. The investment into infrastructure development should obtain the support and endorsement of these groups.

- While the potential scope for HeSANDA is boundless, identification of key types of data, evaluation of data availability, and current clinical trial policies / procedures can inform a phased rollout strategy. To be feasible, HeSANDA should be implemented in stages.

Conclusion

Consistent with the AIHW data development process, the consultations identified and established the business context, information needs, feasibility considerations, and consultation and collaboration requirements for HeSANDA. The process of identifying data for development was also initiated. The consultations established in-principle research community support for HeSANDA and identified a series of principles and recommendations to guide infrastructure development incorporating both distributed ("coordinated") and centralised ("federation") data services as well as stakeholder-endorsed data practices. Additional engagement with the research community as well as key stakeholder groups in clinical trials research should address the remaining steps of the data development process (i.e. developing data elements, field testing, estimating cost of implementation, and obtaining authoritative endorsement) and will help refine the priorities and requirements for a national health studies data asset.