HeSANDA Infrastructure Development

Background to call for registrations of interest

December 2020 (revised February 2021)

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 of data 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. The initial focus of HeSANDA is the establishment of a national infrastructure and standards for the sharing of data from investigator-initiated clinical trials.

Context for HeSANDA Infrastructure Development

The three streams of HeSANDA activity (Data Development, Infrastructure Development, Policy & Culture) work together to:

- Identify the purpose and nature of the national data asset
- Establish the data infrastructure needed to support data sharing and secondary use of data
- Enable buy-in from research communities, patient groups, research institutions, research funders, and health services.

The objective of this coordinated infrastructure initiative is to streamline data sharing for both data producers and secondary users of clinical trials data to maximise the impact of clinical trials data and support research collaboration.

The first phase of Data Development consultations in Q3-4 2020 identified three broad business requirements that HeSANDA shall support. These are:

1. A set of coordinated data services that:
   - Enable access to individual participant-level data (IPD) for secondary use
   - Enable 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 framework
   - Data request and access processes
   - Tools to facilitate data standardisation and compliance

To complement these requirements, the HeSANDA initiative will develop and promote resources (e.g., good practice guidelines, document templates, etc.) for important upstream activities in the research process (such as data capture or participant consent) that can facilitate efficient and effective downstream data sharing. The refinement of these broad requirements and deliverables shall occur via coordinated activities within the three major work streams.

**Overview of Infrastructure Development stream**

The HeSANDA infrastructure will follow a “distributed” model with a network of infrastructure nodes supplying coordinated data services which will be federated via central data discovery, request, and access services. The design and operation of these services will be informed by stakeholder-endorsed coherent data practices. One benefit of the distributed model is that investment in these HeSANDA functions can potentially build upon existing or emerging data infrastructure and trusted research environments operated by health research organisations and support their evolution to a national standard for health research practice.

The Infrastructure Development stream will follow a co-creation model. The 2020 Data Development consultations provided a clear but broad set of business requirements from the clinical trials research community. In the initial co-design phase of the Infrastructure Development stream, participants in the stream will coordinate with the Data Development and Policy & Culture streams to jointly determine HeSANDA’s:
● Information requirements (eg. designing the minimum information requirements and data model)
● Functional requirements (eg. deciding on common data access and request processes)

After these requirements have been established in the co-design phase, participants in the Infrastructure Development stream will implement and deliver infrastructure services to support these requirements at their respective nodes of the network within their own operational contexts (NB. specific software is not mandated but some lightweight common interfaces and information standards will emerge from the requirements co-design phase). This delivery of infrastructure shall follow a co-production approach with ARDC providing co-investment to support the establishment of the infrastructure nodes.

Although the HeSANDA initiative aspires for comprehensive national coverage, this call for participation envisages 10-15 initial infrastructure nodes as a first significant step to enable systematic data sharing for a considerable set of health research data outputs.

Participants will be invited to establish a community governance model for the HeSANDA data asset to oversee and sustain key elements including agreements, technical standards, good practice guidelines, and policies. After the “infrastructure build” phase, ARDC will continue to contribute to the operation of the HeSANDA initiative. As part of the national research infrastructure, ARDC commits to maintaining any central federation services it provides; sustaining the initiative’s governance and broader framework; as well as being open to next steps of evolution such as international integration or expansion of stakeholders beyond clinical trials.

NOTE: As IPD data from clinical trials are sensitive data, there is no expectation that HeSANDA participating services would make such data “open access”. The HeSANDA approach will be guided by the FAIR Data Principles (Findable, Accessible, Interoperable, and Reusable) where the “accessible” principle states that access to data should be as open as possible but as closed as necessary. Mediating access to sensitive data may take into account (for example):

● The Five Safes framework
● Existing community standards of data governance
● The rights of original investigators to be acknowledged and, if appropriate, involved in secondary research

Some of these access arrangements can be enabled by automated technology and infrastructure - some by shared manual business processes. Infrastructure Development participants will co-design the functional requirements of these access arrangements with other key stakeholder groups.

Call for participation

Interest is sought from long-lived organisations in the health research space to establish the HeSANDA infrastructure nodes and co-create the coordinated data services they shall deliver. ARDC envisages an initial set of 10-15 infrastructure nodes that can provide significant national coverage; therefore the Infrastructure Development stream is only open to research networks and consortia. Examples of these include:

● Disease-focused research networks
● Alliances of health research organisations (such as universities, MRI’s, hospitals, etc)
● Health research infrastructure providers with significant regional coverage
The indicators of success for nodes will be:

- Existence of sustainable infrastructure that can be shown to deliver the information and functional requirements agreed to in the co-design phase
- Availability of data from the Nominated Trials (see ‘In-scope activities’ below)
- Capacity and processes to enable data availability for future trials as business as usual

The Infrastructure Development stream will be conducted over a 24 month period from mid 2021 to mid 2023. The initial co-design phase will run for 5 months from July 2021, followed by a 19-month infrastructure build phase that shall conclude in June 2023.

The ARDC proposes to provide co-investment (up to $3m across the nodes) to catalyse and accelerate the emergence of coordinated data services that will underpin the national data asset. Infrastructure Development participants are expected to operate their infrastructure node as part of business as usual after the completion of the Infrastructure Development projects. The ARDC will also provide coordination, digital infrastructure direction, and specialist consultancy services to stream participants. ARDC storage and compute capacity as well as catalogue, identifier, and terminology services are also available to support the HeSANDA program objectives.

Timeline

The HeSANDA Infrastructure Development stream has several phases:

1. **Information session** (February 2021)
2. Registration of interest (March 2021)
3. Facilitation of nodes (April 2021)
4. Submission of proposals (May 2021)
5. Confirmation of node projects (June 2021)
6. Co-design of information and functional requirements (July-November 2021)
7. Distributed infrastructure build phase (December 2021-June 2023)

In-scope activities

This call is for organisations to be HeSANDA infrastructure nodes. HeSANDA nodes will provide the following functions:

- Enable mediated access to clinical trials data
- Hold data from completed clinical trials (unless prohibited by regulatory requirements)
- Provide descriptions of that data (descriptive metadata) using standard
  - structure
  - semantics
  - transfer protocols
- Support common access arrangements to the data
- Support common request and reporting processes

Participating nodes will be required to nominate specific investigator-initiated clinical trials (‘Nominated Trials’) whose data will be made available complying to the HeSANDA framework through the infrastructure developed during the project. It is expected that data from future trials run by participating organisations will by default be made available through their HeSANDA node.
As such, in-scope activities for ARDC co-investment include:

- Co-design of information and functional requirements for HeSANDA nodes
- Infrastructure development to support the functions of a HeSANDA node
- Business process and policy documentation for operation of a HeSANDA node
- Awareness and capability raising of target research communities to support uptake of HeSANDA infrastructure
- Data curation of Nominated Trials
- Automated descriptive metadata creation
- Pipelines and interfaces with existing trial data management systems to enable interoperability with HeSANDA infrastructure

Criteria

The HeSANDA Infrastructure Development stream is a collaborative process. Participants are not competing on distinctive ideas, but rather collaborating together to build a set of compatible nodes within a federated initiative to establish a national standard for health data sharing and research collaboration.

In keeping with the objectives of the NCRIS program, the following criteria will be used to evaluate proposals for co-investment in a HeSANDA node:

- The considerable regional, discipline, or other operational coverage of the participants in any one node
- The volume of health research covered by the participants in a node
- The number of Nominated Trials to be published by the node
- The number of Nominated Trials with funding from the NHMRC (participating funders of the HeSANDA initiative) to be published at the node
- The compelling buy-in from health research communities to publish data from trials via the node on a business as usual footing post-project
- Commitment to work with other nodes on common approaches
- Commitment to sustaining the proposed HeSANDA node and making data from future trials available as part of business-as-usual

ARDC recognises that the COVID-19 situation has placed financial constraints on many organisations, which are restricting the funds available for infrastructure projects. To assist in the continued development of e-research infrastructure, 1:1 matching co-investment will not be mandatory for this open call; however, the level of co-investment will form part of the proposal evaluation criteria.

A portfolio criteria may apply (across all projects) to consider the coverage of all the nodes combined.

Information session

A webinar and Q&A session will be held in February 2021. To attend, please register on our Eventbrite page.