



Help develop a national health studies data asset

29 November 2019

The Australian Research Data Commons ([ARDC](#)) is partnering with the health studies research community to build a distributed national data asset from the outputs of health studies projects. This asset is designed to support further research, meta-analysis, and clinical guideline development.

A first step in that process is a “data development” phase. This phase is to capture initial consensus around:

- data standards and quality appropriate to this purpose, and
- feasible expectations around the nature and accessibility of data “outputs” of health studies projects.

Expressions of interest are sought for this data development phase from:

- Participants
- Contributors / Editors
- Data Development Project Management Organisations.

Program Context

The Health Studies Australian National Data Asset (HeSANDA) program seeks to create a (distributed) data asset from the data outputs of health studies research projects (explicitly but not exclusively from the NHMRC Clinical Trials and Cohort Studies - [CTCS](#) - program). This asset is designed to support further research, meta-analysis, and clinical guideline development.

The three year HeSANDA program does not expect a singular centralised “product”; rather the project will take a first considerable step in that direction by building common standards and practices for project data outputs and a distributed interoperable infrastructure ensuring their availability. The initiative is a step towards health studies data commons, a flagging of strategic intent to transition from “data as a by-product” to “data as an asset”.

In the second half of 2020, consortia of NHMRC Administering Institutions will be invited by the ARDC to opt into projects to build infrastructure that will enable them to contribute to a distributed national data asset with the outputs of their CTCS and other projects (“Infrastructure Development” phase).

The current “Data Development” phase is a public co-design phase preparatory to infrastructure development. In this phase interested parties can help define what the national data asset should look like and what good practice might therefore be for CTCS data outputs. These would serve as requirements and deliverables for the infrastructure building phase.

This is a call for expressions of interest for participation in the initial Data Development phase of the HeSANDA program (phase highlighted in magenta below). This is only a preliminary phase of a multi-streamed program and participation in this phase is **not** a prerequisite for participation in further phases.

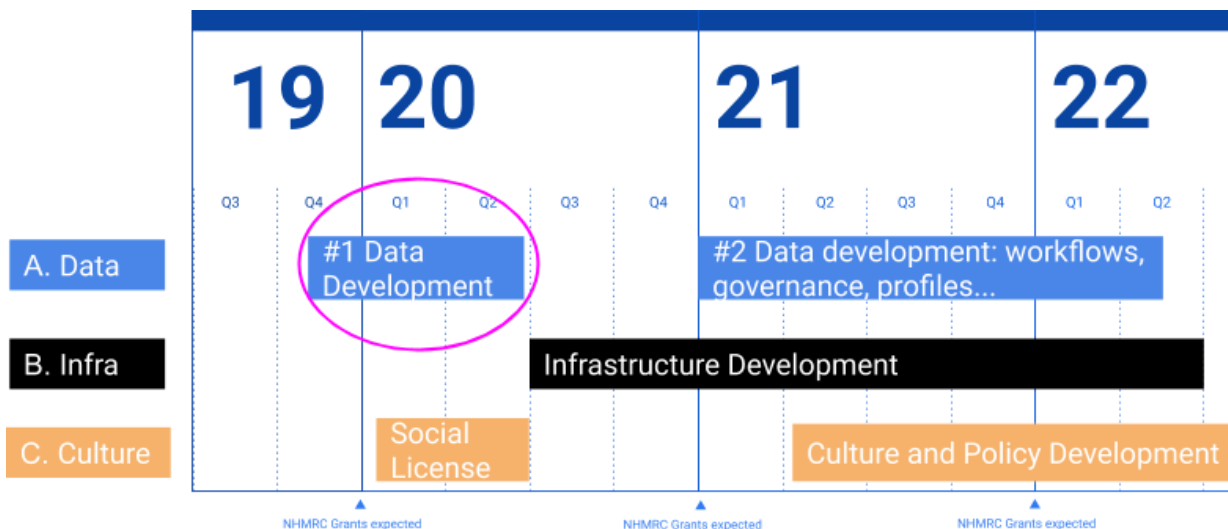


Figure1: HeSANDA program overall indicative schedule

Phase 1: Data Development

Broad Definition of Asset and Project Outputs, Q4 2019 - Q2 2020

The objective of data development¹ is to capture community consensus on:

- the business purpose of the asset (new research, meta analysis, guideline development, integrity etc)
- the kind of data needed to support those purposes
- the appropriate community standards of FAIRness for the asset (Findability, Accessibility, Interoperability, Re-usability)
- mechanisms to enable ongoing community commitment and governance of the asset and its definition

As a first step to building a national health studies data asset, this phase aims to capture rough initial community consensus on the nature of the national data asset. Elements might include:

- Access arrangements (both social and technical)
- Content coverage (individual/aggregate; variables; data structures)
- Semantic quality (initial standard terminologies)

This process is not designed to create a new ISO standard, but rather to identify which existing standards and good practice guidelines are considered by the community to be applicable for the development of a national

¹This process is informed by the AIHW concept of “data development”. See “[A guide to data development](#)” (AIHW)

health studies data asset. This first phase has an aggressive six month timeline which means it envisages a pragmatic set of initial guidelines. An iterative approach is proposed with a second phase of data development in 2021 informed by the beginning of the infrastructure development stream.

ARDC will resource community participation, facilitated discussion, analysis and documentation of a preliminary data asset definition. An important deliverable will be to identify what are appropriate data “outputs” from health studies projects and the feasible pathways from internal project data and processes. A virtual/distributed data asset would be unified by the common practice, standards, accessibility and interoperability identified by this work. “Appropriate” will be a balance of practicality and usefulness. “Fitness for purpose” can only be measured by the agreed business purpose (ie research purpose) of the asset, the documentation of which is itself a fundamental step of this data development phase.

The practical outcome of this initial data development phase will be:

1. an initial consensus on good practice for data outputs of CTCS projects
2. a set of corresponding functional requirements for the infrastructure development phase

Call for Expressions of Interest

ARDC seeks expressions of interest of three types:

1. **Participants** in the data development process:
 - a. Are invited to workshops or webinars
 - b. Are consulted on drafts and proposals
 - c. Provide comments
2. **Editors and authors** of the documentation:
 - a. As above
 - b. Take responsibility for writing sections of and editing the documentation
 - c. Attend fortnightly video conference sessions during Semester 1 2020 for the editorial working group
3. Data development **project management**
 - a. Design the appropriate conceptual framework (with background documentation) for this data development activity
 - b. Manage the scheduling and implementation of activities to successfully deliver phase #1 by July 2020
 - c. Lead and coordinate the editorial workgroup

“Participants” are interested stakeholders (individuals). Expressions of interest are sought so that participants can be registered as members of the project and contacted when appropriate. Participation is on a voluntary basis.

“Editors and authors” are experts (individuals) with experience in documenting data standards and requirements analysis. A modest stipend (a flat \$7K) will be paid to defray time and expenses during the six month project phase and to secure allocation of time. ARDC will select 4-8 “editors/contributors” from the EOI pool to construct an editorial panel with the right spread of expertise and experience.

“Project Manager” is an organisation with an ABN or equivalent who can contract with ARDC to design and manage this data development phase as a project. The organisation should have a national profile and national trust; their expertise in data development and neutral participation should confer confidence in the process. ARDC will proceed to contract with an organisation or organisations from the EOI pool either through direct negotiation or Request for Tender (RfT).

Expressions of interest open on 29 November 2019 should be submitted via the relevant form on the [ARDC website](#) by 8:00pm AEDT **Tuesday 28 January 2020**. If you require further information regarding the process, please [contact ARDC](#).