

Adding Metadata and Data

Once researchers had completed the above steps, the next thing was to get the data they were going to analyse ready on the Workbench.

Researchers were asked to:

- 1 - Upload the meta data for datasets needed in their pilot project to FAIR data services. This is information that describes the datasets and their variables. Researchers were asked to start from the [ICODA Portal](#) to access FAIR Data Services and to review the FAIR online training modules (<https://learn.aridhia.com/>) and the Guide below. If further support was required the Aridhia team were available to help via the service desk at servicedesk@aridhia.com
- 2 - Consider whether they needed support with curation of their datasets. If they required support, projects were asked to contact their project focal point with details so we could plan accordingly.
- 3 - Upload their curated datasets to FAIR data services.

Uploading Data to Project Workspaces - Researchers were asked to:

- 1 - Use the UI or API to upload data to their Project Workspace (see <https://learn.aridhia.com/> for detailed instructions). Once their datasets were in the Workspace, teams could begin the analysis

Uploading metadata to FAIR - a guide

Researchers were asked to use the guide below for uploading their metadata to FAIR. If they required further assistance, they were able to contact the Aridhia service desk at servicedesk@aridhia.com.

[GC ICODA - Metadata Upload to FAIR Guide.pdf](#)

What to include in your Metadata

Dataset Description

Below is an outline of information to include with the dataset description if available.

Dataset information

- A short title which describes the project (without acronyms)
- The purpose that the data was originally collected, such as: Electronic health records, clinical trial data, etc
- The cohorts that the data represents, such as:
 - o Date range cohorts (for example, between 2015-2020)
 - o Geographical cohorts (for example, continents, countries, regions, etc)
 - o Demographic cohorts (for example, age, ethnicity, etc)
 - o Disease type cohorts (for example, HIV, Cardiovascular, COVID-19, etc)
- The type of data that are available, such as:
 - o Level of data (for example, Individual patient level, summary level, etc)
 - o Privacy considerations on data (for example, de-identified, pseudonymised, simulated, etc)
- The estimated size of the dataset
- Links to other datasets details, such as
 - o Code books, data curation or preparation steps, etc

Project information:

- Research question/s that use the dataset
- The project website(s) (can link to the ICODA Driver project website link if useful)
- Links to other project information, such as:
 - o The project protocol, project team, etc
 - o The project outputs, publications, papers, etc

Data access information:

- Information on the organisation or entity that controls the access of this data, including:
 - o Links to how to apply to request access
 - o Dataset access rights (e.g. open access, available for secondary research)
 - o Information on any restrictions on access

Keywords

The more relevant keywords that are used the more likely this dataset will be found by other researchers.

Things to consider when adding keywords:

- Use keywords that you may have used or intend to use within your paper publications
- Use keywords that relate to your research question/s
- Cohort keywords
 - o Geographical cohorts (e.g. continents, countries, regions, etc)
 - o Demographic cohorts (e.g. age, ethnicity, etc)
 - o Disease types (e.g. HIV, Cardiovascular, COVID-19)

- Use synonyms (e.g. child health, children, paediatrics, etc)
- Use variations (e.g. Coronavirus, COVID-19, SARS-CoV-2, etc)

DOI

A digital object identifier (DOI) is a persistent, findable identifier for a resource, in this case the metadata that describes a dataset. This is used to help others find the dataset, and link it to other resources such as research publications, websites, or news articles. If researchers needed help creating a DOI for their metadata, ICODA was able to create one.

Dataset dictionary

The dataset dictionary is the field level details of the dataset tables. The dataset dictionary can contain multiple tables. The fields describe the columns for each table. Each field has the following information:

- Name: the actual name column (is usually shortened, in lowercase, containing no space)
- Label: the full name of the field
- Type: the data type of the field (either Boolean, Integer, Decimal, Text, Date, or Time)
- Description: details of what the field refers to