

# The Ark Documentation.

## Data Set Definition function.

Log in and select your study by clicking the hyperlink corresponding to the name of your study.

Once you have chosen your study we will be able to adjust study details or select modules and functions available in that study (for which you have privileges).

The screenshot shows a web-based application window titled 'prod.the-ark.org.au'. At the top left is the logo for 'Centre for Genetic Epidemiology & Biostatistics, THE UNIVERSITY OF WESTERN AUSTRALIA'. At the top right is the 'THE ARK' logo with the tagline 'OPEN SOURCE SOFTWARE FOR MEDICAL RESEARCH'. The main navigation bar includes 'Study' (which is selected), 'Reporting', and 'Admin'. Below this is a sub-navigation bar with 'Study Details' (selected) and other options like 'Studies', 'Reports', 'Data', 'Analysis', and 'Help'. The 'Study Details' section contains fields for 'ID' (with dropdown values 1, 333, 566, 568, 573, 577, 580, 583), 'Status' (dropdown 'Choose One'), 'Principal Contact' (dropdown), 'Study Name' (input field), and 'Date of Application' (input field with a red calendar icon). Below these are buttons for 'Search', 'Reset', and 'New'. A table below lists study details for each ID:

ID:	Study Name:	Principal Contact:	Date of Application: (dd/mm/yyyy)
1	Lifepool	Lisa Devereux	
333	TIG		
566	Health Twist		
568	PMH Demonstration		
573	PathWest LM Demo		
577	IID Demonstration	Maxine Dennis	
580	LEI		
583	SNU Test		

Alternatively if you have a very long list of studies under your management you may use the search palette/panel at the top of the page. You can search/filter by id, name, principal contact, status of the study or date of application. Once you have added your filtering criteria, click the search button;

The screenshot shows a web-based application interface for managing studies. At the top, there's a header bar with the URL 'prod.the-ark.org.au'. On the left, the 'Centre for Genetic Epidemiology & Biostatistics' logo from the University of Western Australia is visible. On the right, the 'THE ARK' logo with the tagline 'OPEN SOURCE SOFTWARE FOR MEDICAL RESEARCH' is displayed. Below the header, a welcome message 'Welcome arksuperuser@ark.org.au | Logout' is shown. A navigation menu at the top includes 'Study', 'Reporting', and 'Admin'. Under 'Study', a sub-menu 'Study Details' is selected. The main content area contains a search form with fields for 'ID', 'Status' (a dropdown menu), 'Principal Contact' (a text input field), 'Study Name' (with a value 'pedigree' highlighted in blue), and 'Date of Application' (a date input field). Below the search form is a table listing studies. The table has columns: 'ID', 'Study Name', 'Principal Contact', and 'Date of Application (dd/mm/yyyy)'. The data is as follows:

ID:	Study Name:	Principal Contact:	Date of Application: (dd/mm/yyyy)
1	Lifepool	Lisa Devereux	
333	TIG		
566	Health Twist		
568	PMH Demonstration		
573	PathWest LM Demo		
577	IID Demonstration	Maxine Dennis	
580	LEI		
583	SNU Test		
584	random new		
585	IMRC Test		

As you can see the filter has helped us find our chosen study and we can click the hyperlinked study name as discussed before;

Centre for  
Genetic Epidemiology  
& Biostatistics  
THE UNIVERSITY OF WESTERN AUSTRALIA

prod.the-ark.org.au

THE ARK  
OPEN SOURCE SOFTWARE  
FOR MEDICAL RESEARCH

Welcome arksuperuser@ark.org.au | Logout

Study Reporting Admin

Study Details

ID:   
Status: Choose One   
Principal Contact:

Study Name: pedigree  
Date of Application: (dd/mm/yyyy)  17

Search Reset New

ID:	Study Name:	Principal Contact:	Date of Application: (dd/mm/yyyy)
579 << < 1 > >>	Pedigree WIP + Indicates a child study	Dr Adrian Bickerstaffe	07/02/2014

Open # on this page in a new tab the-ark 1.2 Copyright © the-ark.org.au. All rights reserved.

Once you have chosen your study we will be able to adjust study details or select modules and functions available in that study (for which you have privileges). The module we wish to select is Data Sets – then the function “Data Set Definition”.

The data dictionary is the list of longitudinal fields you have to choose from for longitudinal data. Some like to consider this phenotypic data fields, and some have referred to the collections of these fields as datasets or questionnaires. We basically have fields, which can be reused and sets of fields, which can be reused, and re-run over time (ie; longitudinal data).

Over previous Data Set documentation we have created, uploaded, updated the fields in our data dictionary. Now we want to create data sets, aka questionnaires, logical groupings of clinical data, which we may wish to ask once, or hundreds of times over the timeline of our study. This page is where we create those datasets, or (prior to their use) edit them. NOTE: Once a data set has data, it cannot be modified. If we wish to edit an existing data set, we can click on the hyperlinked name of said dataset.

For now we will create a new dataset. Click the “New” button.

The screenshot shows a web-based application titled "prod.the-ark.org.au". At the top, there's a navigation bar with icons for back, forward, search, and refresh, along with the URL. To the right of the URL is a logo for "THE ARK OPEN SOURCE SOFTWARE FOR MEDICAL RESEARCH" featuring a stylized orange and red globe icon. Below the header, a banner for "Centre for Genetic Epidemiology & Biostatistics, THE UNIVERSITY OF WESTERN AUSTRALIA" is visible. The main content area has a title "Study: Pedigree WIP". A horizontal menu bar includes "Study", "Subject", "Datasets", "LIMS", "Reporting", "Work Tracking", "Disease", "Admin", "Data Dictionary", "Data Dictionary Upload", "Data Set Definition", and "Field Data Upload". The "Data Set Definition" tab is currently selected. A sub-menu message says "You may view or edit one of the existing data set definitions or you may create a new data set." with a "New" button. Below this is a table listing existing data set definitions:

ID:	Data Set Name:	Description:	Published:
7	cardiovascular		✓
9	General health check year 2		✓
16	aa	aaa	✓

At the bottom left is a "Display a menu" link, and at the bottom right is a copyright notice: "the-ark 1.2 Copyright © the-ark.org.au. All rights reserved."

Most of these fields and options are similar to what we have seen before. We have a name which we want to meaningfully describe the data set. A description which we should put as much detail as possible.

The list of fields from our data dictionary can be selected from the custom fields selection box. For each field you want in your data set/questionnaire, put it in the selected column by double clicking from the other side, or make your selections and click the arrow to send it in the desired direction. You could also simply upload the names of all fields you want (one field per line) if you have an enormous or external list you wish to use.

The most important field to understand here though is the “published” check box. It is important to understand once the dataset is published, fields can no longer be added or removed in order to ensure the integrity and uniformity of the data set. If you do try to publish a dataset you do at least get a warning/reminder of this **“Custom Fields may not be added or removed once the Data Set is published.”**

The screenshot shows the ARK software interface on a Mac OS X system. The window title is "prod.the-ark.org.au". The header includes the Centre for Genetic Epidemiology & Biostatistics logo, a pedigree tree icon, and THE ARK logo. The top menu bar has items like File, Edit, View, Insert, Tools, Help, and a Log Out button. Below the menu is a navigation bar with tabs: Study, Subject, Datasets, LIMS, Reporting, Work Tracking, Disease, Admin, Data Dictionary (selected), Data Dictionary Upload, Data Set Definition (current tab), and Field Data Upload. The main content area is titled "Study: Pedigree WIP". It contains fields for "Data Set Name" and "Description". A "Custom Fields" section has two lists: "Available" (containing "diasRR", "sysRR", "diasRR\_repeated\_y1", "A", "B") and "Selected" (empty). There is a "Published" checkbox. At the bottom, there's a file input field for "File of Custom Fields" (no file selected), a note about one field per line, and buttons for Save, Cancel, and Delete.

As you can see I have selected 6 fields for a cardiovascular questionnaire/data set. The green text at the top of the page tells me it was successful.

Screenshot of the ARK software interface showing the creation of a new dataset named "TravsCardioSet".

The interface includes the following elements:

- Header: prod.the-ark.org.au, Centre for Genetic Epidemiology & Biostatistics, THE UNIVERSITY OF WESTERN AUSTRALIA, THE ARK OPEN SOURCE SOFTWARE FOR MEDICAL RESEARCH.
- Navigation: Study, Subject, Datasets, LIMS, Reporting, Work Tracking, Disease, Admin.
- Sub-navigation: Data Dictionary, Data Dictionary Upload, Data Set Definition, Field Data Upload. The "Data Set Definition" tab is active.
- Message: "Data Set has been created successfully."
- Form fields:
  - Data Set Name: TravsCardioSet \*
  - Description: Important cardio questions
  - Custom Fields:
    - Available: diasRR\_repeated\_y1, A, B, C
    - Selected: diasRR, sysRR, HeartRate, CHOL, LAST\_DR\_VISIT
  - Published:
  - File of Custom Fields: Choose File (no file selected)
  - Note: One field per line in file. Will replace selected list
- Buttons: Save, Cancel, Delete
- Table: A list of selected fields with columns: ID, Field Name, Question/Label For Field, Sequence, Required, and Allow multiple selections.

ID	Field Name	Question/Label For Field	Sequence	Required	Allow multiple selections
5726	diasRR	diastolic blood pressure	1	X	X
5727	sysRR	systolic blood pressure	2	X	X
5728	HeartRate	Subjects resting heart rate	3	X	X
5729	CHOL		4	X	X
5730	LAST_DR_VISIT		5	X	X
5731	INFARCTION_HISTORY		6	X	X
- Page footer: the-ark 1.2 Copyright © the-ark.org.au. All rights reserved.

Now that you have saved your questionnaire, you can adjust the various fields so say if they are required or not, as seen below

Screenshot of the ARK software interface showing the Data Set Definition page for a dataset named "TravsCardioSet".

The page includes the following sections:

- Data Set Name:** TravscardioSet \*
- Description:** Important cardio questions //
- Custom Fields:** Available Selected
- Published:**
- Custom Field Display Detail:** A modal window showing a successfully updated custom field detail. It includes:
  - ID: 5726
  - Field Name: diasRR
  - Required:
  - Buttons: Save, Cancel
- Note:** One field per line in file. Will replace selected list.
- File of Custom Fields:** Choose File / No file selected. Upload Field List.
- Table:** A grid showing the current list of custom fields. The columns are: ID, Field Name, Question/Label For Field, Sequence, Required, and Allow multiple selections.

ID	Field Name	Question/Label For Field	Sequence	Required	Allow multiple selections
5726	diasRR	diastolic blood pressure	1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
5727	sysRR	systolic blood pressure	2	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
5728	HeartRate	Subjects resting heart rate	3	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
5729	CHOL		4	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
5730	LAST_DR_VISIT		5	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
5731	INFARCTION_HISTORY		6	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
- Navigation:** << < 1 > >>

the-ark 1.2 Copyright © the-ark.org.au. All rights reserved.

It is worth noting the update to the required flag is not uploaded live. This will eventually be fixed but for now, if you exit the panel and re-enter you can see the changes you made.

Screenshot of the ARK software interface showing the Data Set Definition page for a dataset named "TravsCardioSet".

The top navigation bar includes links for Study, Subject, Datasets, LIMS, Reporting, Work Tracking, Disease, Admin, Data Dictionary, Data Dictionary Upload, Data Set Definition (which is currently selected), and Field Data Upload.

The main form fields include:

- Data Set Name: **TravsCardioSet** \*
- Description: **Important cardio questions**
- Custom Fields:
  - Available:  
diasRR\_repeated\_y1  
A  
B  
C
  - Selected:  
diasRR  
sysRR  
HeartRate  
CHOL  
LAST\_DR\_VISIT
- Published:
- File of Custom Fields:  no file selected
- Note: One field per line in file. Will replace selected list
- Buttons: Save, Cancel, Delete

A table below lists the custom fields with their details:

ID:	Field Name:	Question/Label For Field:	Sequence:	Required:	Allow multiple selections:
5726	diasRR	diastolic blood pressure	1	✓	✗
5727	sysRR	systolic blood pressure	2	✗	✗
5728	HeartRate	Subjects resting heart rate	3	✗	✗
5729	CHOL		4	✗	✗
5730	LAST_DR_VISIT		5	✗	✗
5731	INFARCTION_HISTORY		6	✗	✗

Navigation buttons at the bottom left: << < 1 > >>