

# The Ark Documentation.

## Data Centre

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 browser window with the URL `prod.the-ark.org.au`. The page header includes the logo for the Centre for Genetic Epidemiology & Biostatistics at The University of Western Australia, and the logo for THE ARK OPEN SOURCE SOFTWARE FOR MEDICAL RESEARCH. A welcome message reads "Welcome arksuperuser@ark.org.au | Logout". Below the header is a navigation menu with tabs for "Study", "Reporting", and "Admin". The "Study" tab is active, showing a "Study Details" section with a search form. The search form contains fields for "ID:", "Status:" (with a dropdown menu set to "Choose One"), "Principal Contact:", "Study Name:", and "Date of Application:" (with a date picker set to 1/7). There are "Search", "Reset", and "New" buttons. Below the search form is a table of studies.

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	Maxine Dennis	
577	IID Demonstration		
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;

prod.the-ark.org.au

Centre for Genetic Epidemiology & Biostatistics  
THE UNIVERSITY OF WESTERN AUSTRALIA

THE ARK  
OPEN SOURCE SOFTWARE FOR MEDICAL RESEARCH

Welcome arksuperuser@ark.org.au | Logout

Study Reporting Admin

Study Details

ID:  Study Name: pedigree

Status: Choose One Date of Application: (dd/mm/yyyy)  17

Principal Contact:

Search Reset New

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	Maxine Dennis	
577	IID Demonstration		
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;

prod.the-ark.org.au

Centre for Genetic Epidemiology & Biostatistics  
THE UNIVERSITY OF WESTERN AUSTRALIA

THE ARK  
OPEN SOURCE SOFTWARE FOR MEDICAL RESEARCH

Welcome arksuperuser@ark.org.au | Logout

Study Reporting Admin

Study Details

ID:  Study Name: pedigree

Status: Choose One Date of Application: (dd/mm/yyyy)  17

Principal Contact:

Search Reset New

ID:	Study Name:	Principal Contact:	Date of Application: (dd/mm/yyyy)
579	Pedigree WIP	Dr Adrian Bickerstaffe	07/02/2014

<< < > >>

\* indicates a child 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 module we wish to select is Genomics – Data Centre.

Genomics module is created with the intention of allowing collaborating researchers to administer the existing PLINK formatted GWAS datasets and execute GWAS analysis inside High-performance computing sources via a single web user interface.

This simple page is for searching available files and directories stored in the data centers attached to a micro service (These files/directories hold the original datasets to select researchers to start an analysis) and extracting the GWAS information by specific set of filters. As you can see the same standard layout of a list and filterable fields starts us off. To start a file/directory search, first need to select a micro service and then data centre attached to the service by clicking the drop down box.

### **Search for a data source reside in data centre.**

After that clicking the search button, top panel list the directories and files in the root directory of the data centre. By clicking the directory name users can traverse into the directory content and view the sub directories and files. This is an iterative process and users can traverse until the last leaf directory in the file structure. When navigate to specific directory users can use back button to traverse back to its parent directory.

The bottom panel lists the data sources available for selected Micro Service and Data Centre. In addition to above filtering criteria Data Source Id and Data Source Name filters applicable for the data source list. Data Source is the virtual representation of an selected directory in the data centre.

Study: spark

Study Subject Datasets LIMS Reporting Work Tracking Disease Genomics Global Search Admin

Micro Service Data Centre Computation Analysis

Micro Service: spark-local Data Centre: SSH\_HPC

Directory: File Name:

Data Source Id: Data Source Name:

Search Reset

Back	File Name:	Directory:	Absolute File Name:	Directory Status: Unprocessed	Source:
	100000SNP	Yes	/100000SNP	Not Required	Details
	10000SNP	Yes	/10000SNP	Not Required	Details
	1000SNP	Yes	/1000SNP	Not Required	Details
	100SNP	Yes	/100SNP	Not Required	Details
	10SNP	Yes	/10SNP	Not Required	Details
	README.md	No	/README.md	Not Required	Details
	hapmap1	Yes	/hapmap1	Not Required	Details
	1SNP	Yes	/1SNP	Not Required	Details
<< <   > >>					

Data Source Id:	Data Source Name:	Online	Offline	Query
16	10	Online	Offline	Query
21	100	Online	Offline	Query
22	Q22	Online	Offline	Query

## Creating a data source

Click on the detail button of the selected directory will open the details pop-up window. There users can define the existing Type, Owner, Chip, SNP count, and size of the dataset. Click on the Save button will create a new data source which list in the data source panel.

The screenshot shows a web application interface. At the top, there are two dropdown menus: 'Micro Service' set to 'spark-local' and 'Data Centre' set to 'SSH\_HPC'. A 'Details' dialog box is open, containing the following fields:

- Data Source Id:
- Data Source Name:
- Path:
- Type:
- Owner:
- SNP Count:
- Micro Service:
- Description:
- Data Centre:
- Status:
- Chip:
- Size:

At the bottom of the dialog are 'Save', 'Cancel', and 'Delete' buttons. Below the dialog is a table with the following data:

Data Source Id:	Data Source Name:
16	10

At the bottom right of the table are 'Online', 'Offline', and 'Query' buttons.

## Make data source available for querying

Click on the Online button in data source panel will enable data source to be querying. When click the Online button will start a background process to enable dataset ready for querying. To check the online process status users need to refresh the screen by click the Search button. Once the Online process is completed, online button will be disabled and enable the Offline and Query buttons.

## Remove data source from querying

Click on the Offline button in selected data source exist in data source panel which is already in online state. When click the Offline button will start a background process to remove data source from online stage and disable the querying. To check the Offline process status users need to refresh the screen by click the Search button. Once the Offline process is completed, online button will be enabled and disable the Offline and Query buttons.

## Query data source

Click on the Query button in selected data source exist in data source panel which is already in online state. When click the Offline button will open a query Builder pop-up window. There user can select a Id type, and Individual Id. Else selecting a Id type only

will query for all the individuals belongs to the Id type. Other options are Family Id, Gender, Affected Status, and Output type. All the query parameters are optional and without specifying a query parameters will generate a copy of existing dataset. Then click the Query button to start the querying background process and close the pop-up window. There is a new data source is listed in the data source panel after click the Query button.

