

The Ark Documentation.

Data Dictionary Upload 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 links for 'Study', 'Reporting', and 'Admin'. The 'Study' link is currently selected. Below the navigation is a sub-menu titled 'Study Details'. The main content area contains form fields for 'ID' (with value 1), 'Status' (dropdown menu 'Choose One'), 'Principal Contact' (text input), 'Study Name' (text input), and 'Date of Application' (text input with a calendar icon). Below these fields are buttons for 'Search', 'Reset', and 'New'. A table below lists study details for various IDs:

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 Dictionary Upload”.

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).

From our previous work, the example study below has fields based on systolic and diastolic blood pressure and a few other poorly named fields. These could be used in multiple collections of questions that will be repeated over time. If we want a convenient way to add many more fields at once, we can upload a CSV file containing the information about the fields rather than manually enter a field per page as we did in the Data Dictionary documentation.

The screenshot shows the ARK software interface. At the top, there is a navigation bar with icons for back, forward, search, and other functions. The URL 'prod.the-ark.org.au' is displayed. On the left, there is a logo for 'Centre for Genetic Epidemiology & Biostatistics, THE UNIVERSITY OF WESTERN AUSTRALIA'. In the center, there is a decorative graphic of a tree with people around it. On the right, there is a logo for 'THE ARK OPEN SOURCE SOFTWARE FOR MEDICAL RESEARCH' with a stylized orange sphere.

Welcome [arksuperuser@ark.org.au](#) | [Logout](#)

Study: Pedigree WIP

Study Subject Datasets LIMS Reporting Work Tracking Disease Admin

Data Dictionary Data Dictionary Upload Data Set Definition Field Data Upload

ID: Field Name:
Field Type: Description:
Units:
Minimum Value: Maximum Value:

ID:	Field Name:	Field Type:	Description:	Question/Label For Field:	Units:	Min:	Max:
319	A	NUMBER	a		kg		
320	B	DATE	b				
321	C	CHARACTER	c				
211	diasRR	NUMBER	diastolic blood pressure	diastolic blood pressure			
216	diasRR_repeated_y1	NUMBER		diasRR Slightly Different Question?			
212	sysRR	NUMBER	systolic blood pressure	systolic blood pressure			

<< < 1 >>

We will now add a 8th, 9th and 10th field Via an upload all at once.

After clicking Data Dictionary Upload, the following will appear, containing previous uploads for that study (with their results, completion status, the uploaded file itself and a report). A Download Template button also exists which will download the template of the data you need to upload;

The screenshot shows theARK software interface. At the top, there are browser-style navigation buttons (back, forward, search, etc.) and a URL bar showing "prod.the-ark.org.au". Below the URL bar are three logos: "Centre for Genetic Epidemiology & Biostatistics" (University of Western Australia), a stylized tree icon, and "THE ARK OPEN SOURCE SOFTWARE FOR MEDICAL RESEARCH". The main header says "Welcome arksuperuser@ark.org.au | Logout" and "Study: Pedigree WIP". A navigation menu below the header includes "Study", "Subject", "Datasets", "LIMS", "Reporting", "Work Tracking", "Disease", "Admin", "Data Dictionary" (which is highlighted in blue), "Data Dictionary Upload", "Data Set Definition", and "Field Data Upload".

Step 1/5: Select data file to upload

Select the file containing data, the file type and the specified delimiter, click Next to continue.

Filename: no file selected * Delimiter Type: *

Note: When uploading an Excel file (.xls), delimiter type is ignored. However, we recommend converting (.xls to a CSV file) and examining the data prior to uploading. The variability of xl version means we cannot be certain of the quality of data the XL file will represent

Note: Click "Download Template" for an example file. Header columns are required.

Next Cancel Finish

Upload ID:	Filename:	File Format:	UserID:	Start Time:	Finish Time:	Download Template	Status:
149	DataDictionaryUpload.csv	CSV	demo@ark.org.au	2014-09-26 17:16:40.0	2014-09-26 17:16:40.0	<input type="button" value="Download File"/> <input type="button" value="Download Report"/>	Successfully completed

<< < 1 > >>

Display a menu

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

As described in the Data Dictionary documentation, some fields are not so self-explanatory and I will explain those below. The ID field is a system based uneditable field. Field Type can currently only be a number, text, or date. Number fields can have min/max values, and these limits will be strictly enforced on upload of data. A default value will be filled in if empty values are uploaded on a data set upload.

One field that confused some people is the encoded value field. This is described further in the custom field uploader documents. But basically, the idea is (for convenience or to enable data migrations/uploads from another system) you can upload shorthand values like 0 for no, 1 for yes by setting up encoded values as follows "0=no;1=yes;" (minus the quotes).

Utilizing the template, I have then added the following to a .CSV file.

	A	B	C	D	E	F	G	H	I
1	FIELD_NAME	FIELD_TYPE	DESCRIPTION	QUESTION	UNITS	ENCODED_V	MINIMUM_V	MAXIMUM_V	MISSING_VAL
2	CHOL	NUMBER	a				0	25	
3	LAST_DR_VISIT	DATE	b						
4	INFARCTION_HISTORY	CHARACTER	c			1=Yes;2=No;			
5									
6									
7									

Once you have added the data you wish in a similar manner, then click save and name the file something logical which will be easy to remember and look back at;

Click the Choose File button, then select the CSV document you saved in the previous step. It is safest to always use the COMMA delimiter (CSV) in the Delimiter Type option.

Then click the Next button;

The screenshot shows a web-based application window titled "prod.the-ark.org.au". At the top left is the "Centre for Genetic Epidemiology & Biostatistics" logo from The University of Western Australia. To its right is a stylized tree icon. On the far right is the "THE ARK" logo with the tagline "OPEN SOURCE SOFTWARE FOR MEDICAL RESEARCH". The main menu bar includes "Study", "Subject", "Datasets", "LIMS", "Reporting", "Work Tracking", "Disease", and "Admin". Below this is a secondary navigation bar with "Data Dictionary", "Data Dictionary Upload" (which is highlighted in blue), "Data Set Definition", and "Field Data Upload".

Step 1/5: Select data file to upload

Select the file containing data, the file type and the specified delimiter, click Next to continue.

Filename: * Delimiter Type:

Note: When uploading an Excel file (.xls), delimiter type is ignored. However, we recommend converting (.xls to a CSV file) and examining the data prior to uploading. The variability of xl version means we cannot be certain of the quality of data the XL file will represent

Note: Click "Download Template" for an example file. Header columns are required.

Next Cancel Finish

Upload ID:	Filename:	File Format:	UserID:	Start Time:	Finish Time:	Download Template	Status:
149	DataDictionaryUpload.csv	CSV	demo@ark.org.au	2014-09-26 17:16:40.0	2014-09-26 17:16:40.0	<input type="button" value="Download File"/> <input type="button" value="Download Report"/>	Successfully completed

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

If you have no error messages your data will display as below. If you do have error messages, they will describe what is necessary to fix, and then you must go back to step one after having fixed your CSV file. If you work from the template, you will typically be just fine. If you have no errors and your code has been parsed as expected, click "Next".



Welcome [arksuperuser@ark.org.au](#) | [Logout](#)

Study: Pedigree WIP

Study Subject Datasets LIMS Reporting Work Tracking Disease Admin

Data Dictionary Data Dictionary Upload Data Set Definition Field Data Upload

Step 2/5: File Format Validation

The file format has been validated. If there are no errors, click Next to continue.

Validation Messages:

Data Preview: (First 20 rows of file)

ROW	FIELD_NAME	FIELD_TYPE	DESCRIPTION	QUESTION	UNITS	ENCODED_VALUES	MINIMUM_VALUE	MAXIMUM_VALUE	MISSIN
1	A	NUMBER	a		kg				999
2	B	DATE	b					01/01/	
3	C	CHARACTER	c			1=Yes;2=No;			999

[Next](#) [Cancel](#) [Finish](#)

[Display a menu](#)

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

At this point data itself is validated, in this instance I didn't Save my CSV correctly and so it had told me I will be updating existing fields! If I had intended to do this, I simply click the "update existing rows" checkbox and it will update those fields after clicking next.



Welcome [arksuperuser@ark.org.au](#) | [Logout](#)

Study: Pedigree WIP

Study Subject Datasets LIMS Reporting Work Tracking Disease Admin

Data Dictionary Data Dictionary Upload Data Set Definition Field Data Upload

Step 3/5: Data Validation

The data in the file is now validated, correct any errors and try again, otherwise, click Next to continue.

Validation Messages:

Data on row 1 exists, please confirm update
Data on row 2 exists, please confirm update
Data on row 3 exists, please confirm update

[Download Validation Messages](#)

Update existing data? —

Data Preview: (First 20 rows of file)

ROW	FIELD_NAME	FIELD_TYPE	DESCRIPTION	QUESTION	UNITS	ENCODED_VALUES	MINIMUM_VALUE	MAXIMUM_VALUE	MISSING_VALUE
1	A	NUMBER	a		kg				999
2	B	DATE	b						01/01/3000
3	C	CHARACTER	c			1=Yes;2=No;			999

KEY: Indicates data in error Indicates data with a warning Indicates data to insert Indicates data to update

[Next](#) [Cancel](#) [Finish](#)

However that wasn't my intention so I corrected the file, did step 1,2 all over again. Now at this point, If you have no error messages your data will display as below (all the green looks good, the data is what I expected, I can continue). If you do have error messages, they will describe what is necessary to fix (like on the previous page), and then you must go back to step one after having fixed your CSV file. If you work from the template, you will typically be just fine. If you have no errors and your code has been parsed as expected, click "Next".



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
Subject
Datasets
LIMS
Reporting
Work Tracking
Disease
Admin

Data Dictionary
Data Dictionary Upload
Data Set Definition
Field Data Upload

Step 3/5: Data Validation

The data in the file is now validated, correct any errors and try again, otherwise, click Next to continue.

Validation Messages:

Data Preview: (First 20 rows of file)

ROW	FIELD_NAME	FIELD_TYPE	DESCRIPTION	QUESTION	UNITS	ENCODED_VALUES	MINIMUM_VALUE	MAXIMUM_VALUE	MISSING_VALUE
1	CHOL	NUMBER	cholesterol				0	25	
2	LAST_DR_VISIT	DATE	last time you visited Dr						
3	INFARCTION_HISTORY	CHARACTER	Any past MIs			1=Yes;2=No;			

KEY: Indicates data in error Indicates data with a warning Indicates data to insert Indicates data to update

[Next](#) [Cancel](#) [Finish](#)

[Display a menu](#)

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

If that all worked fine, you get the following step to confirm the changes. Click next.

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". In the center is a decorative graphic of a tree with many colorful circular leaves. To the right is the "THE ARK" logo with the tagline "OPEN SOURCE SOFTWARE FOR MEDICAL RESEARCH". The main menu bar includes "Study", "Subject", "Datasets", "LIMS", "Reporting", "Work Tracking", "Disease", and "Admin". Below this is a secondary navigation bar with "Data Dictionary", "Data Dictionary Upload", "Data Set Definition", and "Field Data Upload". The main content area is titled "Step 4/5: Confirm Upload" and contains the text: "Data will now be written to the database, click Next to continue, otherwise click Cancel." At the bottom of this area are three buttons: "Next", "Cancel", and "Finish".

Display a menu

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

You have now confirmed changes, step 5 should show up which is just a confirmation from The Ark that your upload is SUBMITTED. That does not actually mean it is completed as the message hints (since fixed in our code base). Rather, a background job has been submitted. It will be complete when you come back to Data Dictionary upload later and see a status of "Successfully Completed"

The screenshot shows a web browser window for 'prod.the-ark.org.au'. The header includes the Centre for Genetic Epidemiology & Biostatistics logo (a blue stylized 'S' icon) and the THE ARK logo (a red 'ARK' with a yellow/orange circular graphic). The top navigation bar has links for Study, Subject, Datasets, LIMS, Reporting, Work Tracking, Disease, Admin, Data Dictionary, Data Dictionary Upload, Data Set Definition, and Field Data Upload. A sub-menu for 'Data Dictionary Upload' is open. A central message box displays the text: 'Step 5/5: Data Upload Finished' and 'The data has been successfully submitted, When the actual upload is finished it will have a status of "Successfully Completed".' Below the message are 'Next', 'Cancel', and 'Finish' buttons. At the bottom left is a 'Display a menu' link, and at the bottom right is copyright information: 'the-ark 1.2 Copyright © the-ark.org.au. All rights reserved.'

See the first row is the upload we just did, the status says successfully completed. You can check the report to see if results were as expected

Screenshot of the Ark software interface showing the Data Dictionary Upload step.

The browser address bar shows: prod.the-ark.org.au

The header includes the Centre for Genetic Epidemiology & Biostatistics logo (University of Western Australia) and THE ARK logo (Open Source Software for Medical Research).

Welcome arksuperuser@ark.org.au | Logout

Study: Pedigree WIP

Menu tabs: Study, Subject, Datasets, LIMS, Reporting, Work Tracking, Disease, Admin

Sub-menu tabs: Data Dictionary, Data Dictionary Upload, Data Set Definition, Field Data Upload

Message: Data upload of file: DataDictionaryUpload.csv was uploaded successfully

Step 1/5: Select data file to upload

Select the file containing data, the file type and the specified delimiter, click Next to continue.

Filename: Choose File no file selected * Delimiter Type: COMMA *

Note: When uploading an Excel file (.xls), delimiter type is ignored. However, we recommend converting (.xls to a CSV file) and examining the data prior to uploading. The variability of xl version means we cannot be certain of the quality of data the XL file will represent.

Note: Click "Download Template" for an example file. Header columns are required. Next Cancel Finish

Upload ID:	Filename:	File Format:	User ID:	Start Time:	Finish Time:	Download Template	Status:
703	DataDictionaryUpload.csv	CSV	arksuperuser@ark.org.au	2015-12-16 17:56:05.0	2015-12-16 17:56:05.0	Download File Download Report	Successfully completed
149	DataDictionaryUpload.csv	CSV	demo@ark.org.au	2014-09-26 17:16:40.0	2014-09-26 17:16:40.0	Download File Download Report	Successfully completed

<< < 1 > >>

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