

The Ark Documentation.

How to use the subject uploader to create a new subject.

Instruction Path a (with perfect data and no validation)

Log in and select our study.

Once we have chosen our study we will see the following (Study) functions, the last of which is Subject Upload. Click that function.

The screenshot shows the Ark software interface on a Mac OS X desktop. The window title is "Welcome to The Ark". The URL in the address bar is "prod.the-ark.org.au/ark/home?3". The interface includes the Centre for Genetic Epidemiology & Biostatistics logo (University of Western Australia), the lifepool logo, and the THE ARK logo (Open Source Software for Medical Research).

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

Study: Lifepool

Study Subject Datasets LIMS Reporting Work Tracking Disease Admin

Study Details Study Components Manage Users Subject Custom Fields Subject Custom Field Upload Subject Upload

Study:	Lifepool	Summary of study statistics:	Total Subjects: 23405
ID:	1	Study Name:	Lifepool *
Description:	lifepool is an initiative of the National Breast Cancer	Status:	Active *
Date of Application:	(dd/mm/yyyy)	Estimated Year of Completion:	
Chief Investigator:	ian Campbell *	Co-Investigator:	
Contact Person:	Lisa Devereux	Contact Phone:	9656 1096
Application Modules: *	<u>Available</u> <u>Selected</u>	Logo:	<input type="file"/> Choose File No file chosen
Disease		Max 100px in height, & less than 100Kb in size	
Auto-generate Subject UID: <input type="checkbox"/> Note: Disabled if Subjects exist	Subject UID Pattern: <input type="text"/> [Prefix][Token][Counter (padded # chars)]	Auto-consent subjects: <input type="checkbox"/>	Start Subject counter at: <input type="text"/> e.g. 1, 1000 or 5000
Auto-generate Biocollection UID: <input type="checkbox"/> Note: Disabled if Biocollections exist		Auto-generate Biospecimen UID: <input type="checkbox"/> Note: Disabled if Biospecimens exist	

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

The following page will appear. You will see the following sections from top to bottom;

1. The files upload section. This is where you can begin the steps to upload subjects.
2. The Labels and Template download section. This will be described further below.
3. The Result list section. This section lists all previous uploads.



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

Study: Lifepool

[Study](#) [Subject](#) [Datasets](#) [LIMS](#) [Reporting](#) [Work Tracking](#) [Disease](#) [Admin](#)

[Study Details](#) [Study Components](#) [Manage Users](#) [Subject Custom Fields](#) [Subject Custom Field Upload](#) [Subject Upload](#)

Step 1/5: Select data file to upload
Select the file containing data, upload type and the specified delimiter, click Next to continue.

Filename: No file chosen * Upload Type: *

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](#)

ID:	Filename:	File Format:	User ID:	Start Time:	Finish Time:	Download Subject Template	Download Custom Field Template	Download Consent Field Template	Status:
224	SUBJECTDEMOGRAPHICSLifePoolINT.csv	CSV	arksuperuser@ark.org.au	2014-10-26 02:47:32.0	2014-10-26 02:47:56.0	Download File	Download Report		Successfully completed
223	SUBJECTDEMOGRAPHICSLifePoolINT.csv	CSV	arksuperuser@ark.org.au	2014-10-26 02:28:43.0	2014-10-26 02:29:06.0	Download File	Download Report		Successfully completed

The next step is to upload your data. You have two choices for your starting point;

1. You can select Download Subject Template if you are unsure which fields to include. The template lists which fields are mandatory, what type/format of data is expected and valid values for fields that are limited in that way (see below).

A	B	C	D	E	F	G	H	I
1	SUBJECTUID	TITLE						
2	DESCRIPTION	The unique identifier	The title by which the subject prefers to be addressed					
3	FORMAT	CHARACTER	CHARACTER					
4	MANDATO	Yes - unless assign	No					
5	VALID VAL	Unknown, Br, Capt, Col, Cpl, Dean, Dr, Fr, Lac, Major, Miss						
6	NOTE:	Rer						
7								
8								

Once you have added your data to the template remove rows 2-6, so that your data now begins on row 2. Also remove column A, so that SUBJECTUID is the first column.

2. Alternatively, you can access an example from the usefulTools/csvTemplatesForUploaders directory of our code repository. If you have not been given a local version of our repository you can access it from

<https://code.google.com/p/ark-informatics/source/browse/#svn%2FARK%2Ftrunk%2FusefulTools%2FcsvDataForUploaders>
For the example below I utilized the SubjectUploadFromWhite.csv template file.

Once you have taken the steps listed above to ensure you have the fields you wish to upload, save the file, then you simply;

1. Select the filename of that file in the “The files upload section”.
2. Select the “Subject Demographic Data” upload type
3. Leave the delimiter type as CSV (even if using an excel spreadsheet).
4. Click the Next button once.

The screenshot shows a web-based application window for the ARK software. The top navigation bar includes the ARK logo and the URL prod.the-ark.org.au/ark/home?3. The left sidebar shows a file tree with categories like All My Files, Applications, Documents, Downloads, db-scripts, support, migration, tendersby, ark-informatics, usefulTools, Macintosh HD, Movies, and Music. A specific file, 'SubjectUploadFromWhite.csv', is highlighted in blue. The main content area displays a table of uploaded files:

ID:	Filename:	File Format:	User ID:	Start Time:	Finish Time:	Download Subject Template	Download Custom Field Template	Download Consent Field Template	Status:
224	SUBJECTDEMOGRAPHICSLifePoolINT.csv	CSV	arksuperuser@ark.org.au	2014-10-26 02:47:32.0	2014-10-26 02:47:56.0	Download File	Download Report		Successfully completed
223	SUBJECTDEMOGRAPHICSLifePoolINT.csv	CSV	arksuperuser@ark.org.au	2014-10-26 02:28:43.0	2014-10-26 02:29:06.0	Download File	Download Report		Successfully completed

This will take you to the step 2. This page will let you know if the file structure is correct for the upload type. For example, if I misnamed the address columns (ADDRESS_LINE_1 and ADDRESS_LINE_2 instead of "STREET_ADDRESS" and "BUILDING_NAME" as the template and uploader expected) I would get the following type of error message;

The screenshot shows a web-based application for managing medical research data. At the top, there are navigation links for 'Study', 'Subject', 'Datasets', 'LIMS', 'Reporting', 'Work Tracking', 'Disease', and 'Admin'. Below these, a sub-navigation bar includes 'Study Details', 'Study Components', 'Manage Users', 'Subject Custom Fields', 'Subject Custom Field Upload' (which is highlighted), and 'Subject Upload'. The main content area is titled 'Step 2/5: File Format Validation' and contains the following text: 'The file format has been validated. If there are no errors, click Next to continue.' A 'Validation Messages:' section lists three errors: 'Error: The specified file does not appear to conform to the expected file format. Please refer to the template, as seen on step one, for the correct format.', 'Error: the column name ADDRESS_LINE_1 is not a valid column name.', and 'Error: the column name ADDRESS_LINE_2 is not a valid column name.'. Below this, a 'Data Preview: (First 20 rows of file)' table is shown, containing the following data:

ROW	SUBJECTUID	TITLE	FIRST_NAME	MIDDLE_NAME	LAST_NAME	PREFERRED_NAME	DATE_OF_BIRTH	VITAL_STATUS	GENDER	STATUS	DATE_OF_DEATH	CAUSE_OF_DEATH
1	998877	Mr	Paul	Andrew	White	Paul	4/02/1963	Alive	Male	Subject		

At the bottom of the page are 'Next', 'Cancel', and 'Finish' buttons.

We can look back at the template and see that this is the expected naming, and correct our mistake.

R	S
BUILDING_NAME	STREET_ADDRESS
BUILDING_NAME	STREET_ADDRESS
CHARACTER	CHARACTER
No	No

We click cancel at the bottom of the page and start the upload again

So now that I have corrected the headers on my file, I repeat step one and see if there is any more mistakes

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

Study: Lifepool

Study Subject Datasets LIMS Reporting Work Tracking Disease Admin

Study Details Study Components Manage Users Subject Custom Fields Subject Custom Field Upload Subject 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	SUBJECTUID	TITLE	FIRST_NAME	MIDDLE_NAME	LAST_NAME	PREFERRED_NAME	DATE_OF_BIRTH	VITAL_STATUS	GENDER	STATUS	DATE_OF_DEATH	CAUSE_OF_DEATH
1	998877	Mr	Paul	Andrew	White	Paul	4/02/63	Alive	Male	Subject		

Next Cancel Finish

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

SubjectUpload (1).xls Show All

I see no validation messages at step 2, so I can assume everything is ok and click next at the bottom of the page (this step will validate all of your **data** Vs what is expected, so can take some time on very large files. Please be patient and don't click back, as it is busy doing work and will let you know when it is complete).

We are now at step 3. This step shows if anything went wrong in step 2. The validation messages say I made an incorrect assumption of fields for SILENT, it should have a Boolean (true or false) type. So I have to fix the data and try again.

Welcome to The Ark 0% 20% off Living Social (Max) Travis

prod.the-ark.org.au/ark/home?3

Centre for Genetic Epidemiology & Biostatistics THE UNIVERSITY OF WESTERN AUSTRALIA

lifepool

THE ARK OPEN SOURCE SOFTWARE FOR MEDICAL RESEARCH

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

Study: Lifepool

Study Subject Datasets LIMS Reporting Work Tracking Disease Admin

Study Details Study Components Manage Users Subject Custom Fields Subject Custom Field Upload Subject 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:

Error: Row 1: Subject UID: 998877 SILENT: SILENT is not in the valid boolean value. Please use true or false for this column.

Download Validation Messages

Data Preview: (First 20 rows of file)

ROW	SUBJECTUID	TITLE	FIRST_NAME	MIDDLE_NAME	LAST_NAME	PREFERRED_NAME	DATE_OF_BIRTH	VITAL_STATUS	GENDER	STATUS	DATE_OF_DEATH	CAUSE_OF_DEATH
1	998877	Mr	Paul	Andrew	White	Paul	4/02/63	Alive	Male	Subject		

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

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

SubjectUpload (1).xls

If we have corrected all of our data error we will now see an empty box for the validation messages.

prod.the-ark.org.au/ark/home?3

Centre for Genetic Epidemiology & Biostatistics THE UNIVERSITY OF WESTERN AUSTRALIA

lifepool

THE ARK OPEN SOURCE SOFTWARE FOR MEDICAL RESEARCH

Welcome arksuperuser@ark.org.au | Logout

Study: Lifepool

Study Subject Datasets LIMS Reporting Work Tracking Disease Admin

Study Details Study Components Manage Users Subject Custom Fields Subject Custom Field Upload Subject 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	SUBJECTID	TITLE	FIRST_NAME	MIDDLE_NAME	LAST_NAME	PREFERRED_NAME	DATE_OF_BIRTH	VITAL_STATUS	GENDER	STATUS	DATE_OF_DEATH	CAUSE_OF_DEATH
1	998877	Mr	Paul	Andrew	White	Paul	4/02/63	Alive	Male	Subject		

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

Next Cancel Finish

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

SubjectUpload (1).xls Show All

You can therefore click next if you are happy with what you see.

(note: this upload can also update demographic data if you so desire, in which case the uploader will highlight those rows that are to be updated. This is generally considered a sub-optimal way to manage data by going off line and then merging back with the online database, as you risk losing the state that was maintained online).

You have now confirmed you are accepting the data, click next one more time and a BATCH process will begin.

The screenshot shows the Ark software interface for a study named 'Lifepool'. At the top, there are navigation links for 'Study', 'Subject', 'Datasets', 'LIMS', 'Reporting', 'Work Tracking', 'Disease', and 'Admin'. Below these, a sub-navigation bar includes 'Study Details', 'Study Components', 'Manage Users', 'Subject Custom Fields', 'Subject Custom Field Upload' (which is highlighted), and 'Subject Upload'. A central message box displays the text 'Step 4/5: Confirm Upload' and 'Data will now be written to the database, click Next to continue, otherwise click Cancel.' At the bottom of the message box are three buttons: 'Next', 'Cancel', and 'Finish'.

You now have completed the final step and can click finish whenever you want...

The screenshot shows the Ark software interface for a study named 'Lifepool'. The layout is identical to the previous screenshot, with the same navigation and sub-navigation bars. 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".' At the bottom of the message box are three buttons: 'Next', 'Cancel', and 'Finish'.

BUT must know that the data has been submitted BUT it has not been completed until the result list page says that your upload is in a “Successfully Completed” status.

This is what a successful upload will look like on the result list section. You can also see what you two previous unsuccessful attempts looked like (the first time we failed validation of the file (headers or file type/format), the second time we failed validation of the data.

The screenshot shows the Ark software interface. At the top, there are logos for the Centre for Genetic Epidemiology & Biostatistics, Lifepool, and The ARK. The URL in the browser is prod.the-ark.org.au/ark/home?3. The user is logged in as arksuperuser@ark.org.au. The main menu includes Study, Subject, Datasets, LIMS, Reporting, Work Tracking, Disease, Admin, and a dropdown for Study Details, Study Components, Manage Users, Subject Custom Fields, Subject Custom Field Upload, and Subject Upload. A success message 'Data upload of file: SubjectUploadFromWhite.csv was uploaded successfully' is displayed. Below this, a step-by-step guide for Step 1/5: Select data file to upload is shown, along with fields for Filename (Choose File, No file chosen) and Upload Type (Choose One). The Delimiter Type is set to COMMA. A note about Excel file handling is present. At the bottom are Next, Cancel, and Finish buttons. The results list table shows three entries:

ID	Filename	File Format	User ID	Start Time	Finish Time	Status
532	SubjectUploadFromWhite.csv	CSV	arksuperuser@ark.org.au	2015-03-23 13:16:51.0	2015-03-23 13:16:51.0	Download Subject Template Download Custom Field Template Download Consent Field Template Download Pedigree Template Successfully completed
531	SubjectUploadFromWhite.csv	CSV	arksuperuser@ark.org.au	2015-03-23 13:13:12.0		Download File Error while validating data
530	SubjectUploadFromWhite.csv	CSV	arksuperuser@ark.org.au	2015-03-23 13:10:56.0		Download File Error validation file

As you can see, in the results list for our successful upload, we have the options of Download File (this will download the original file you used) or Download Report which will give you a summary of what was uploaded, and any assumptions (or less serious errors) that were made upon successful rows. See an example below;

The screenshot shows the Ark software interface with a detailed log report titled 'uploadReport532.txt'. The log contains the following information:

```

*****
the-ark.org.au Upload Report
Mon Mar 23 13:16:50 EST 2015
*****
Study: Lifepool
UserID: arksuperuser@ark.org.au
Filename: SubjectUploadFromWhite.csv
File Format: CSV
Delimiter Type: COMMA
Warning/Info: could not find a state named 'WA' in Australia for row 1, but will proceed.
Info: Using the default status 'Current' and the default type 'RESIDENTIAL' for row 1, but will proceed.
Total file size: 0.00 MB
Processed 1 rows.
Inserted 1 subjects.
Updated 0 subjects.

```

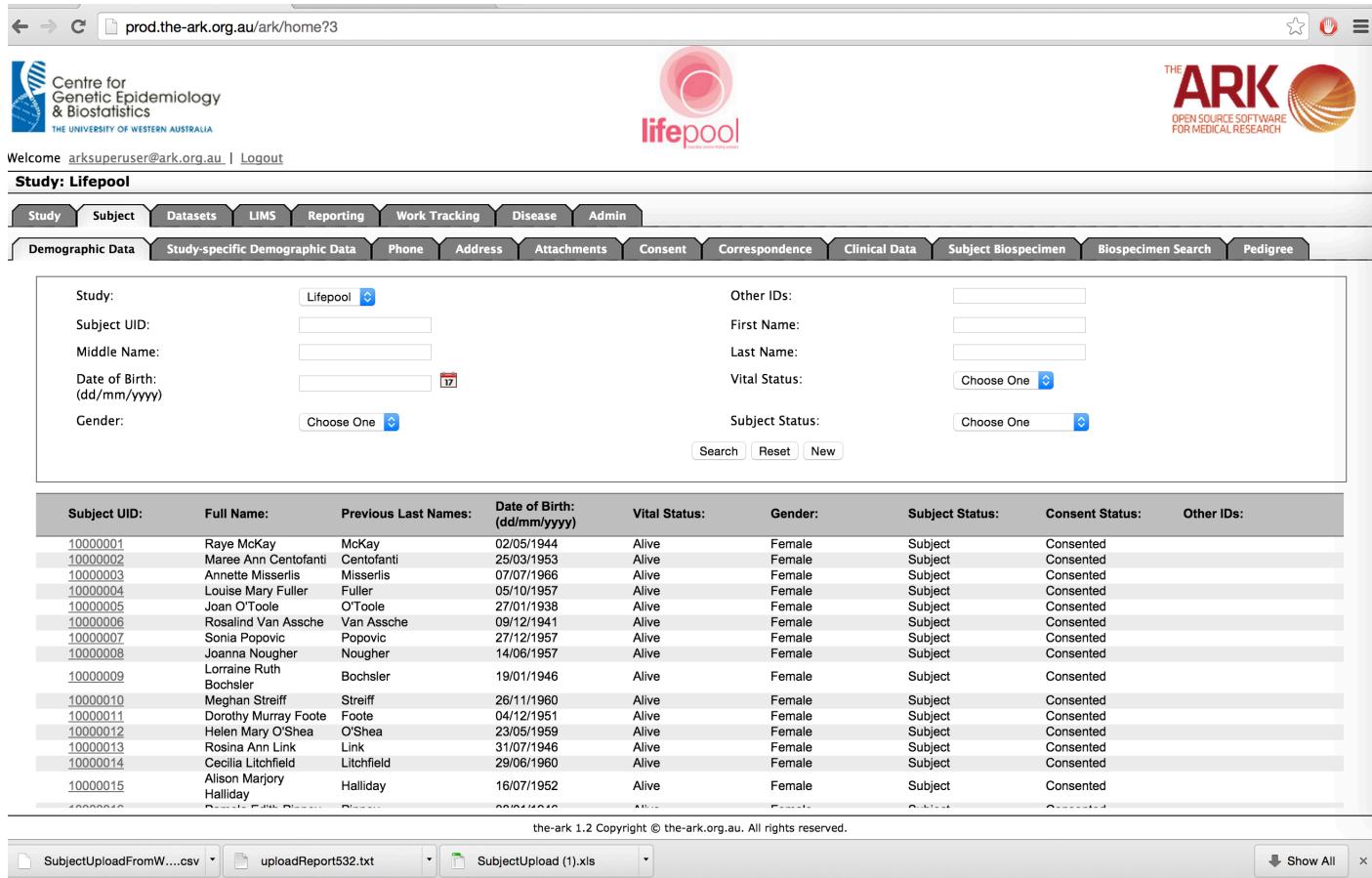
The log ends with a status message: Successfully completed.

The results list table shows three entries:

ID	Filename	File Format	User ID	Start Time	Finish Time	Status
224	SUBJECTDEMOGRAPHICSLifePoolINT.csv	CSV	arksuperuser@ark.org.au	2014-10-26 02:47:32.0	2014-10-26 02:47:56.0	Download File Download Report Successfully completed
223	SUBJECTDEMOGRAPHICSLifePoolINT.csv	CSV	arksuperuser@ark.org.au	2014-10-26 02:28:43.0	2014-10-26 02:29:06.0	Download File Download Report Successfully completed
222	SUBJECTDEMOGRAPHICSLifePoolINT.csv	CSV	arksuperuser@ark.org.au	2014-10-26	2014-10-26	Download File Download Report Successfully completed

As always, we recommend you test what you just did to ensure the results were as you expected.

To do this, we click on the Subject Module Tab near the top of the page.



Welcome arksuperuser@ark.org.au | Logout

Study: Lifepool

Study Subject Datasets LIMS Reporting Work Tracking Disease Admin

Demographic Data Study-specific Demographic Data Phone Address Attachments Consent Correspondence Clinical Data Subject Biospecimen Biospecimen Search Pedigree

Study: Lifepool
Subject UID:
Middle Name:
Date of Birth:
(dd/mm/yyyy)
Gender:
Other IDs:
First Name:
Last Name:
Vital Status:
Choose One
Subject Status:
Choose One

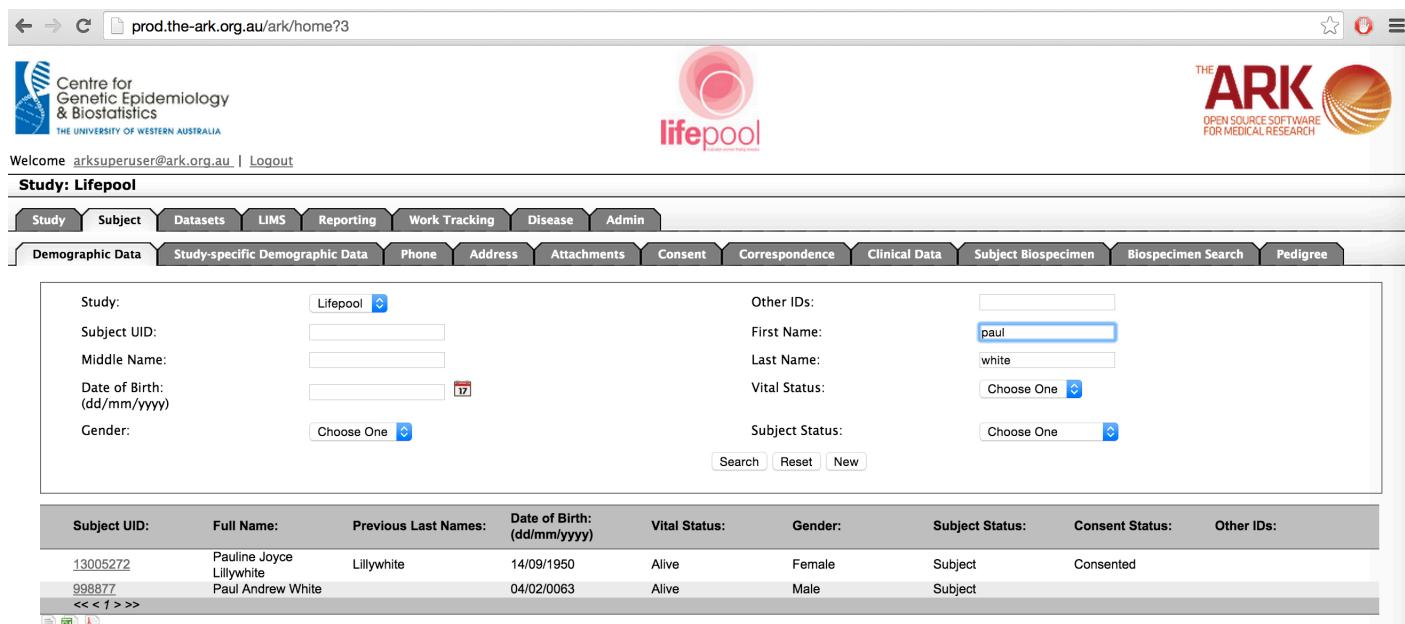
Search Reset New

Subject UID:	Full Name:	Previous Last Names:	Date of Birth: (dd/mm/yyyy)	Vital Status:	Gender:	Subject Status:	Consent Status:	Other IDs:
10000001	Raye McKay	McKay	02/05/1944	Alive	Female	Subject	Consented	
10000002	Maree Ann Centofanti	Centofanti	25/03/1953	Alive	Female	Subject	Consented	
10000003	Annette Misserlis	Misserlis	07/07/1966	Alive	Female	Subject	Consented	
10000004	Louise Mary Fuller	Fuller	05/10/1957	Alive	Female	Subject	Consented	
10000005	Joan O'Toole	O'Toole	27/01/1938	Alive	Female	Subject	Consented	
10000006	Rosalind Van Assche	Van Assche	09/12/1941	Alive	Female	Subject	Consented	
10000007	Sonia Popovic	Popovic	27/12/1957	Alive	Female	Subject	Consented	
10000008	Joanna Nougher	Nougher	14/06/1957	Alive	Female	Subject	Consented	
10000009	Lorraine Ruth Bochsler	Bochsler	19/01/1946	Alive	Female	Subject	Consented	
10000010	Meghan Streiff	Streiff	26/11/1960	Alive	Female	Subject	Consented	
10000011	Dorothy Murray Foote	Foote	04/12/1951	Alive	Female	Subject	Consented	
10000012	Helen Mary O'Shea	O'Shea	23/05/1959	Alive	Female	Subject	Consented	
10000013	Rosina Ann Link	Link	31/07/1946	Alive	Female	Subject	Consented	
10000014	Cecilia Litchfield	Litchfield	29/06/1960	Alive	Female	Subject	Consented	
10000015	Alison Marjory Halliday	Halliday	16/07/1952	Alive	Female	Subject	Consented	
10000016	Ryanne Farnie	Farnie	02/04/1960	Alive	Female	Subject	Consented	

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

SubjectUploadFromW....csv uploadReport532.txt SubjectUpload (1).xls Show All

We can use a search term to find out subjects we wish to test;



Welcome arksuperuser@ark.org.au | Logout

Study: Lifepool

Study Subject Datasets LIMS Reporting Work Tracking Disease Admin

Demographic Data Study-specific Demographic Data Phone Address Attachments Consent Correspondence Clinical Data Subject Biospecimen Biospecimen Search Pedigree

Study: Lifepool
Subject UID:
Middle Name:
Date of Birth:
(dd/mm/yyyy)
Gender:
Other IDs:
First Name:
paul
Last Name:
white
Vital Status:
Choose One
Subject Status:
Choose One

Search Reset New

Subject UID:	Full Name:	Previous Last Names:	Date of Birth: (dd/mm/yyyy)	Vital Status:	Gender:	Subject Status:	Consent Status:	Other IDs:
13005272	Pauline Joyce Lillywhite	Lillywhite	14/09/1950	Alive	Female	Subject	Consented	
998877	Paul Andrew White		04/02/0063	Alive	Male	Subject	Consented	

If you have come this far, congratulations. As you can see we found the name we just uploaded (Paul White). Click on his SubjectUID to test his uploaded data.

Hopefully all the data you expected is there. "Paul White" is now in context and you can go ahead and click on any information that you expect to see from him.

prod.the-ark.org.au/ark/home?3

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

Study: Lifepool Subject UID: 998877

Study **Subject** Datasets LIMS Reporting Work Tracking Disease Admin

Demographic Data Study-specific Demographic Data Phone Address Attachments Consent Correspondence Clinical Data Subject Biospecimen Biospecimen Search Pedigree

Subject UID: 998877 * Title: Mr

Other IDs: Source: Other ID

First Name: Paul Middle Name: Andrew
Last Name: White Preferred Name: Paul
Note: Updated when Last Name changed

Date of Birth: 04/02/0063 Vital Status: Alive

Date Last Known Alive:

Date of Death:

Gender: Male Subject Status: Subject *

Marital Status: Married Preferred Contact Method: Choose One

Email: paul.white@uwa.edu.au Other Email:

Email Status: Unknown Email Status: (other): Unknown

Heard About Study From:

Comments:

Consent Date: Consent Status: Choose One *