

The Ark Documentation.

How to use the subject custom field uploader to insert custom field data.

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 header includes the Centre for Genetic Epidemiology & Biostatistics logo, the lifepool logo, and the THE ARK OPEN SOURCE SOFTWARE FOR MEDICAL RESEARCH logo. The main menu bar has "Travis" selected. Below the menu is a navigation bar with links for "Study", "Subject", "Datasets", "LIMS", "Reporting", "Work Tracking", "Disease", and "Admin". A secondary navigation bar below it includes "Study Details", "Study Components", "Manage Users", "Subject Custom Fields", "Subject Custom Field Upload", and "Subject Upload". The "Subject Custom Field Upload" tab is currently active. The main content area displays the "Study Details" for the "Lifepool" study. It shows the study name "Lifepool", a summary of "Total Subjects: 23405", and various configuration fields. These include "ID: 1", "Description: lifepool is an initiative of the National Breast Cancer", "Date of Application: dd/mm/yyyy (17)", "Chief Investigator: Ian Campbell *", "Contact Person: Lisa Devereux", "Application Modules: * Available (checkboxes for Disease, Study, Subject, Datasets, Genotypic, LIMS) Selected (checkboxes for Study, Subject, Datasets, Genotypic, LIMS)", "Logo: Max 100px in height, & less than 100Kb in size (choose file button, no file chosen, yellow info icon)", and "Auto-generate Subject UID: [checkbox] Note: Disabled if Subjects exist". There are also sections for "Auto-consent subjects: [checkbox]", "Start Subject counter at: [text input field]", and "Auto-generate Biospecimen UID: [checkbox]". At the bottom of the form, there is a note: "Note: Disabled if Biospecimens are disabled". The footer of the window says "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.

←1

←2

←3

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

1. You can select Download Custom Field 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
SUBJECTUID	YOUR_FIRST_CUSTOM_FIELD	YOUR_SECOND_CUSTOM_FIELD	AND SO ON	
DESCRIPTION: The unique identifier assigned for this subject	Value for first custom field	Value for second field	AND SO ON	
NOTE: Rer				

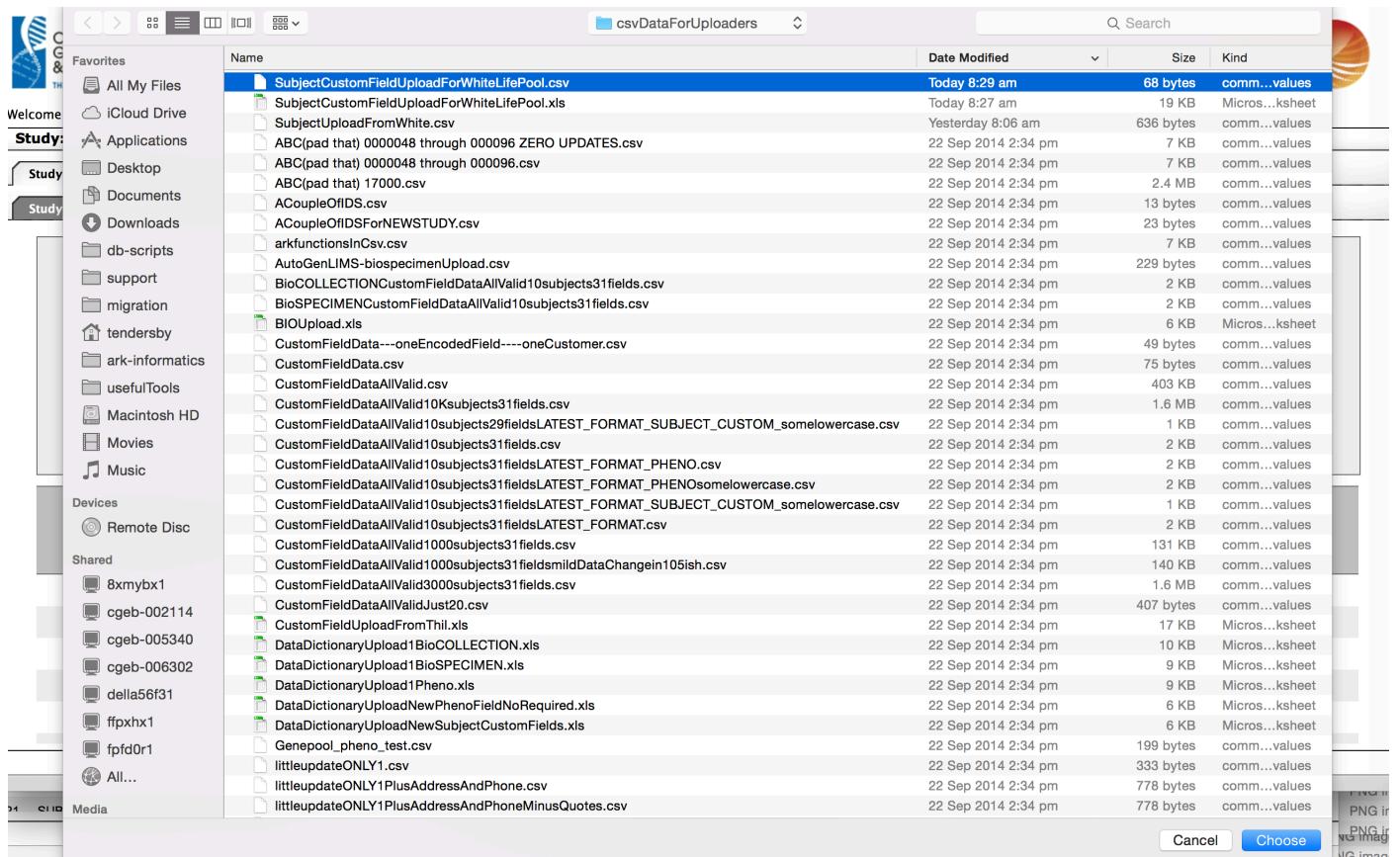
As you can see, this template is more of a guide than a generated file. You have to fill out the field names in the headers eg;

A	B	C
1 SUBJECTUID	CANCER_BREAST	CANCER_DX_DATE
2	998877 Unknown	1/01/01

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 (Created) the SubjectCustomFieldUploadForWhiteLifePool.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 “Study-specific (custom) Data” upload type
3. Leave the delimiter type as CSV (even if using an excel spreadsheet).
4. Click the Next button once.



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 had included a field in the headers, which didn't exist, I would get the following type of error message;

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:

The specified delimiter type was: ,.
This file is not valid because:
The column name does not match with an existing study-specific custom field.

Download Validation Messages

Data Preview: (First 20 rows of file)

ROW	SUBJECTUID	CANCER_BREAST	CANCER_DX_DATE
1	998877	Unknown	1/01/01
2			

Next Cancel Finish

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

This one turns out to be a rather cryptic issue due to the fact Excel had included an extra comma in its conversion to a CSV file. We can see this by using a normal text editing tool to view the CSV as a text file.

```
IRDConc26Feb2015.sql *
SubjectCustomFieldUploadForWhiteLifePool.csv *

1 SUBJECTUID,CANCER_BREAST,CANCER_DX_DATE,
2 998877,Unknown,1/01/01,
3 ,,,
```

I could manually manipulate the file and remove the excess comma that is on the end of every line, or use a regular expression replacement if I were aware of how to do that. The best approach might be to learn how to make this NOT happen in Excel (see next page).

Ensuring Excel doesn't add extra commas and lines;

Step 1: Select all (or at least those that may have once had data) columns after our last column of actual data

A	B	C	D	E	F
SUBJECTUID	CANCER_BREAST	CANCER_DX_DATE			
998877	Unknown	1/01/01			
3					
4					
5					
6					
7					
8					
9					
10					
11					

Step 2: Delete selected columns

A	B	C	D	E	F	G
SUBJECTUID	CANCER_BREAST	CANCER_DX_DATE				
998877	Unknown	1/01/01				
3						
4						
5						
6						
7						
8						
9						
10						
11						

Step 3 & 4; Repeat the process by selecting and deleting all rows after our last row of data.

A screenshot of Microsoft Excel showing a context menu open over a row of data. The menu includes options such as Cut (⌘X), Copy (⌘C), Paste (⌘V), Paste Special... (^⌘V), Insert, and Delete. The data in the cells includes SUBJECTUID, CANCER_BREAST, and CANCER_DX_DATE. The Excel ribbon is visible at the top, and the formula bar shows A3.

	A	B	C	D	E	F	G
1	SUBJECTUID	CANCER_BREAST	CANCER_DX_DATE				
2	998877	Unknown	1/01/01				
3							
4							
5							
6							
7							
8							
9							
10							
11							

Step 5: Now save as CSV file.

A screenshot of the 'Save As' dialog in Mac OS X. The 'Format:' dropdown is set to 'Comma Separated Values (.csv)'. The 'Description' section explains that the data is exported to a text file using commas to separate values. There are buttons for 'Options...', 'Compatibility Report...', and 'Save'. The 'Cancel' button is also visible.

Format: Comma Separated Values (.csv)

Description

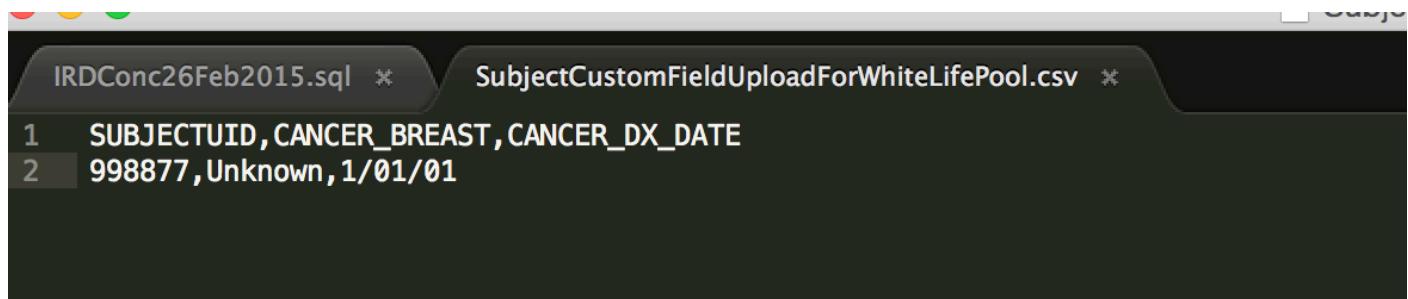
Exports the data on the active sheet to a text file that uses commas to separate values in cells.

Learn more about file formats

Options... Compatibility Report... ⚠ Compatibility check recommended

New Folder Cancel Save

Step 6: Check your data in a plain text editor (eg; Notepad, sublime edit, vi, vim,TextEdit). It should now have no extra commas and no extra lines.



```
IRDConc26Feb2015.sql *
SubjectCustomFieldUploadForWhiteLifePool.csv *
1 SUBJECTUID,CANCER_BREAST,CANCER_DX_DATE
2 998877,Unknown,1/01/01
```

So now our upload data is as expected, we can try again.

We click cancel at the bottom of The Ark 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

The screenshot shows the Ark software interface. At the top left is the Centre for Genetic Epidemiology & Biostatistics logo from The University of Western Australia. In the center is the lifepool logo. At the top right is THE ARK logo with the tagline "OPEN SOURCE SOFTWARE FOR MEDICAL RESEARCH". The top navigation bar includes links for Study, Subject, Datasets, LIMS, Reporting, Work Tracking, Disease, Admin, Study Details, Study Components, Manage Users, Subject Custom Fields, Subject Custom Field Upload, and Subject Upload. The current page is Step 2/5: File Format Validation for the Lifepool study. A message states: "The file format has been validated. If there are no errors, click Next to continue." Below this is a "Validation Messages" section which is currently empty. A "Data Preview: (First 20 rows of file)" section shows the following table:

ROW	SUBJECTID	CANCER_BREAST	CANCER_DX_DATE
1	998877	Unknown	1/01/01

At the bottom of the page are buttons for Next, Cancel, and Finish. A footer note reads: "the-ark 1.2 Copyright © the-ark.org.au. All rights reserved."

I see no validation messages at step 2, so I can assume everything is ok and click next at the bottom of the page (the next 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 Unknown is not a valid encrypted value, as it is expecting the key value, ie; 9. So I have to fix the data and try again.

The screenshot shows the Ark software interface. At the top, there are three logos: Centre for Genetic Epidemiology & Biostatistics (University of Western Australia), lifepool, and THE ARK OPEN SOURCE SOFTWARE FOR MEDICAL RESEARCH. The main navigation bar includes links for Study, Subject, Datasets, LIMS, Reporting, Work Tracking, Disease, Admin, Study Details, Study Components, Manage Users, Subject Custom Fields, Subject Custom Field Upload, and Subject Upload. The current page is 'Subject Custom Field Upload'. A sub-section titled 'Step 3/5: Data Validation' displays a validation message: 'The data in the file is now validated, correct any errors and try again, otherwise, click Next to continue.' Below this, a 'Validation Messages' section contains the error: 'Subject UID: 998877: The field CANCER_BREAST value Unknown is not in the expected encoded values: 1=Yes;2=No;9=Unknown;'. There is also a link 'Download Validation Messages'. At the bottom left, there is a checkbox for 'Continue despite the fact data has failed validation?' with the note 'Note: Data will be flagged in error and should be fixed manually.' A 'Data Preview' table shows the first 20 rows of the file, with columns for ROW, SUBJECTUID, CANCER_BREAST, and CANCER_DX_DATE. The first row is shown as: 1, 998877, Unknown, 1/01/01. The footer of the page reads 'the-ark 1.2 Copyright © the-ark.org.au. All rights reserved.'

Click cancel and go and fix your data, and then repeat steps one and two again.

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

The screenshot shows the ARK software interface for the 'Lifepool' study. At the top, there are three logos: 'Centre for Genetic Epidemiology & Biostatistics' (University of Western Australia), 'lifepool', and 'THE ARK OPEN SOURCE SOFTWARE FOR MEDICAL RESEARCH'. The main navigation bar includes links for Study, Subject, Datasets, LIMS, Reporting, Work Tracking, Disease, Admin, Study Details, Study Components, Manage Users, Subject Custom Fields, Subject Custom Field Upload (which is highlighted in yellow), and Subject Upload. A sub-navigation bar under 'Subject Custom Field Upload' lists 'Study 3/5: Data Validation'. Below this, a message states: 'The data in the file is now validated, correct any errors and try again, otherwise, click Next to continue.' A section titled 'Validation Messages:' contains an empty box. At the bottom, a 'Data Preview: (First 20 rows of file)' table is shown:

ROW	SUBJECTID	CANCER_BREAST	CANCER_DX_DATE
1	998877	9	1/01/01

At the very bottom of the page, a footer bar reads: 'the-ark 1.2 Copyright © the-ark.org.au. All rights reserved.'

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

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 in red), and 'Subject Upload'. A large central box 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 right of this 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 large central box is titled 'Step 5/5: Data Upload Finished' and contains the text: 'The data has been successfully submitted, When the actual upload is finished it will have a status of "Successfully Completed".' At the bottom right of this 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.

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

o Data upload of file: SubjectCustomFieldUploadForWhiteLifePool.csv was uploaded successfully

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

Filename: Choose File no file selected * Upload Type: Choose One *

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

ID:	Filename:	File Format:	User ID:	Start Time:	Finish Time:	Status:
537	SubjectCustomFieldUploadForWhiteLifePool.csv	CSV	arksuperuser@ark.org.au	2015-03-26 12:51:45.0	2015-03-26 12:51:45.0	Successfully completed
536	SubjectCustomFieldUploadForWhiteLifePool.csv	CSV	arksuperuser@ark.org.au	2015-03-26 12:15:18.0	2015-03-26 12:15:18.0	Error while validating data
535	SubjectCustomFieldUploadForWhiteLifePool.xls	XLS	arksuperuser@ark.org.au	2015-03-26 11:51:27.0	2015-03-26 11:51:27.0	Error validation file
534	SubjectCustomFieldUploadForWhiteLifePool.csv	CSV	arksuperuser@ark.org.au	2015-03-26 11:50:31.0	2015-03-26 11:50:31.0	Error validation file
533	SubjectUploadFromWhite.csv	CSV	arksuperuser@ark.org.au	2015-03-24 18:50:49.0	2015-03-24 18:50:49.0	Successfully completed
				2015-03-23	2015-03-23	

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

As you can see, in the results list for our successful upload, we have the options of Download File (this will down 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 theARK web application interface. At the top, there are three logos: 'Centre for Genetic Epidemiology & Biostatistics' (University of Western Australia), 'lifepool', and 'THE ARI OPEN SOURCE SOF FOR MEDICAL RESE'. The main area has a 'Welcome arksuperuser@ark.org.au | Logout' message.

Study: the-ark.org.au Upload Report
Thu Mar 26 12:51:44 EST 2015

Study: Lifepool
UserID: arksuperuser@ark.org.au
Filename: SubjectCustomFieldUploadForWhiteLifePool.csv
File Format: CSV
Delimiter Type: COMMA
Total file size: 0.00 MB
Inserted 1 rows of data
2 fields were inserted.
0 fields were updated.

Subject Upload

Upload Type: Choose One *

Examine the data prior to uploading.

Start Cancel Finish

Time:	Finish Time:	Status:
Download Subject Template	Download Custom Field Template	Download Consent Field Template
Download Pedigree Template		
537 SubjectCustomFieldUploadForWhiteLifePool.csv	CSV arksuperuser@ark.org.au 2015-03-26 12:51:45.0	Download File Download Report Successfully completed
536 SubjectCustomFieldUploadForWhiteLifePool.csv	CSV arksuperuser@ark.org.au 2015-03-26 12:15:18.0	Download File Error while validating data
535 SubjectCustomFieldUploadForWhiteLifePool.xls	XLS arksuperuser@ark.org.au 2015-03-26 11:51:27.0	Download File Error validation file
534 SubjectCustomFieldUploadForWhiteLifePool.csv	CSV arksuperuser@ark.org.au 2015-03-26 11:50:31.0	Download File Error validation file
	2015-03-24 2015-03-24	

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

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

Demographic Data

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 Nouher	Nouher	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	Pauline Joyce Lillywhite	Lillywhite	02/09/1960	Alive	Female	Subject	Consented	

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

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

Welcome arksuperuser@ark.org.au | Logout

Study: Lifepool

Demographic Data

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		

As you can see we found the name we just uploaded (Paul White). Click on his SubjectUID to test put this subject in Context.

Your next step is to click Study-specific Demographic Data to see if the data was uploaded as expected. Hopefully all the data you expected is there. As we can see, our two fields are there as we expected.

The screenshot shows the Ark software interface for managing study data. At the top, there are navigation links for 'Study', 'Subject', 'Datasets', 'LIMS', 'Reporting', 'Work Tracking', 'Disease', and 'Admin'. The current page is 'Demographic Data' for Subject UID 998877, under the 'Study-specific Demographic Data' tab. The main content area contains the following fields:

- Send subject a newsletter? (dropdown menu: Choose One)
- Medicare number (text input field)
- Medicare dataset consent (dropdown menu: Choose One)
- Subject diagnosed with breast cancer? (dropdown menu: Unknown)
- Breast cancer diagnosis date (text input field: 01/01/0001)
- Source of breast cancer diagnosis (dropdown menu: Choose One)
- Other source of breast cancer diagnosis (text input field)
- Surgeon providing treatment of breast cancer (text input field)
- Clinic of the surgeon providing treatment (text input field)
- Address of the surgical clinic (text input field)
- Surgery comments (text input field)
- Hospital providing treatment (text input field)
- Address of the hospital (text input field)
- Doctor providing treatment of breast cancer (text input field)
- Clinic of the doctor providing treatment (text input field)
- Address of the doctor (text input field)
- Patient type (dropdown menu: Choose One)
- Pathology lab (text input field)

One thing to note is that we entered the data in DD/MM/YY format. The Ark took this literally as 01/01/0001, if we wanted to enter that as 01/01/2001 we need to enter it exactly that way "01/01/2001". We will do that in our next guide "The Ark Documentation – SubjectCustomFieldUploader – Updating"