

The Ark Documentation.

How to use the subject custom field uploader to UPDATE 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.

Screenshot of the Ark Study Management interface showing the 'Subject Custom Field Upload' tab selected.

The interface includes the following sections:

- Study Details:** Lifepool
- Study Components:** Available: Disease; Selected: Study, Subject, Datasets, Genotypic, LIMS
- Subject Custom Fields:** (not visible in screenshot)
- Subject Custom Field Upload:** (selected tab)
- Subject Upload:** (not visible in screenshot)

Form fields (Study Details tab):

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: *	Available: Disease	Selected: Study, Subject, Datasets, Genotypic, LIMS	Logo: Max 100px in height, & less than 100Kb in size <input type="file"/> Choose File No file chosen

Form fields (Subject Custom Field Upload tab):

Auto-generate Subject UID: <input type="checkbox"/> Note: Disabled if Subjects exist	Auto-consent subjects: <input type="checkbox"/>
Subject UID Pattern: <input type="text"/> [Prefix][Token][Counter (padded # chars)]	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 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.

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

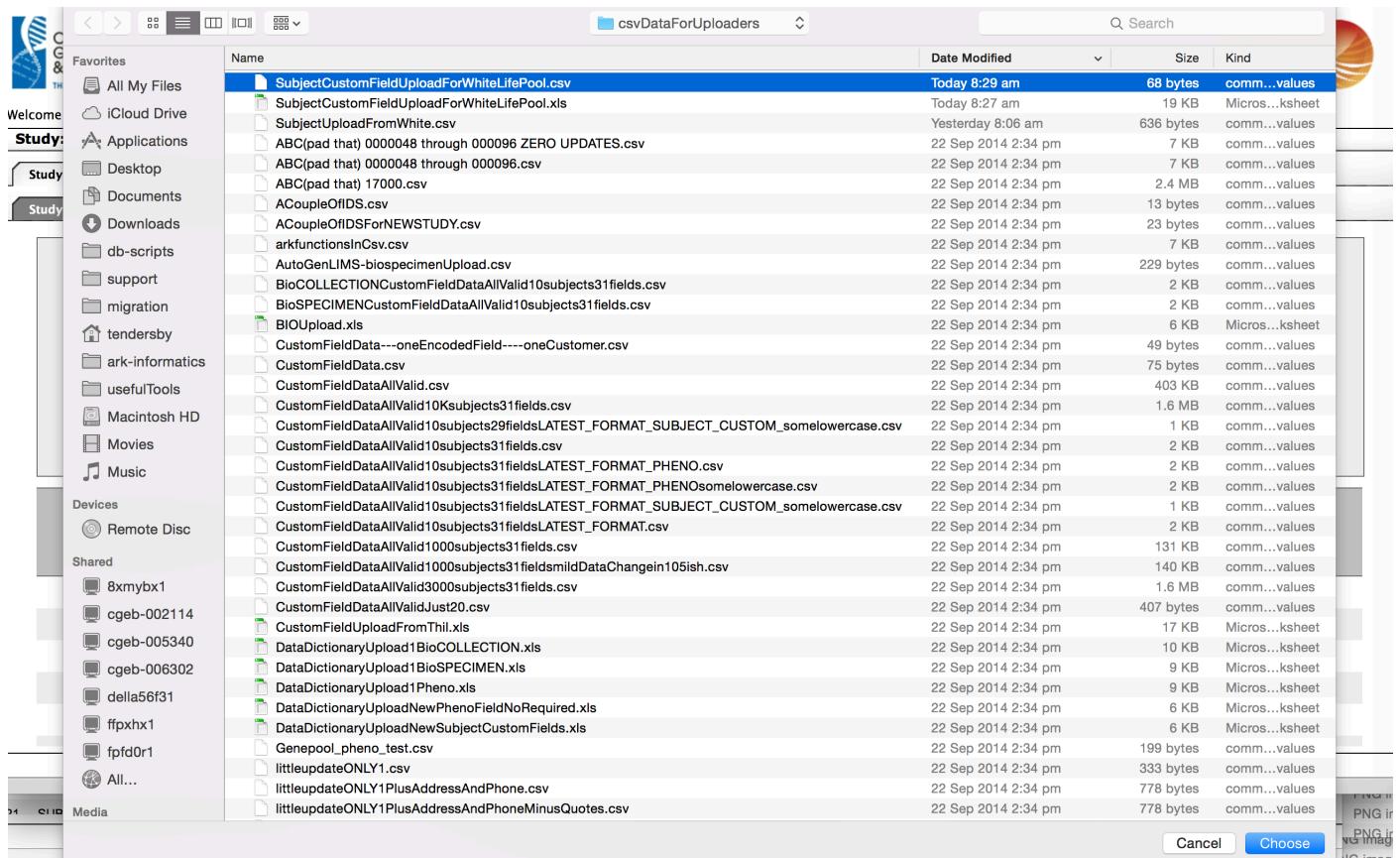
2. Alternatively, you can access an example from the usefulTools/csvTemplatesForUploaders directory of our code repository. If you have not been given a local version of our repository you can access it from

<https://code.google.com/p/ark-informatics/source/browse/#svn%2FARK%2Ftrunk%2FusefulTools%2FcsvDataForUploaders>

For the example below I utilized (updated) the SubjectCustomFieldUploadForWhiteLifePool.csv template file. You could also use the file you originally used to insert the data, but I strongly advise that you ONLY use the data you intend to update in order to avoid losing the current state of your data.

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
1	SUBJECTUID	CANCER_BREAST	CANCER_DX_DATE			
2	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
1	SUBJECTUID	CANCER_BREAST	CANCER_DX_DATE				
2	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 options are:

- Cut
- Copy
- Paste
- Paste Special...
- Insert
- Delete

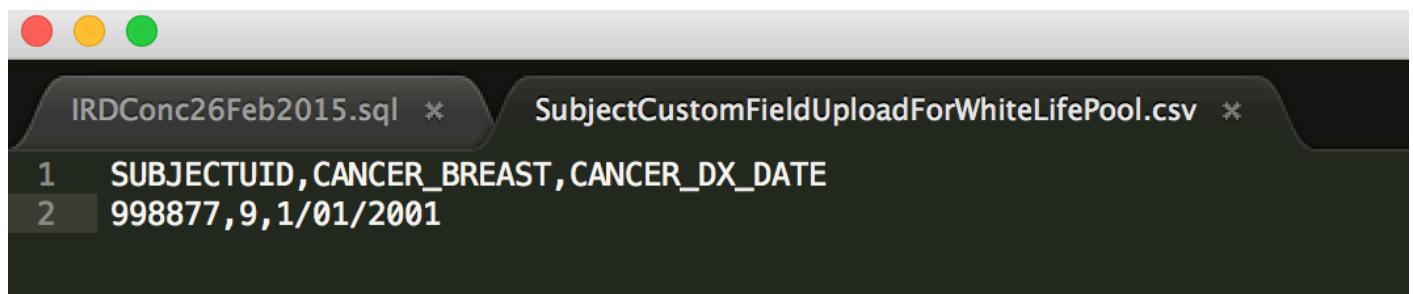
The menu also includes keyboard shortcuts: ⌘X, ⌘C, ⌘V, ⌘V, ⌘V, and ⌘V respectively. The background shows a spreadsheet with columns A through G and rows 1 through 11. Row 1 contains "SUBJECTUID", "CANCER_BREAST", and "CANCER_DX_DATE". Row 2 contains "998877", "Unknown", and "1/01/01". Rows 3 through 10 are empty. Row 11 contains the context menu options.

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 states: 'Exports the data on the active sheet to a text file that uses commas to separate values in cells.' There are buttons for 'Options...', 'Compatibility Report...', and 'Save'. A warning message at the bottom right says '⚠ Compatibility check recommended'.

Name	Date Modified	Size	Kind
SubjectCu...ifePool.csv	Today 9:05 am	62 bytes	comm...vali
SubjectCu...ifePool.xls	Today 9:05 am	18 KB	Micros...ksl
SubjectUp...White.csv	Yesterday 8:06 am	636 bytes	comm...vali
ABC(pad t...DATES.csv	22 Sep 2014 2:34 pm	7 KB	comm...vali
ABC(pad t...000096.csv	22 Sep 2014 2:34 pm	7 KB	comm...vali
ABC(pad that) 17000.csv	22 Sep 2014 2:34 pm	2.4 MB	comm...vali
ACoupleOfIDS.csv	22 Sep 2014 2:34 pm	13 bytes	comm...vali
ACoupleOf...TUDY.csv	22 Sep 2014 2:34 pm	23 bytes	comm...vali
arkfunctionsInCsv.csv	22 Sep 2014 2:34 pm	7 KB	comm...vali
AutoGenLI...Upload.csv	22 Sep 2014 2:34 pm	229 bytes	comm...vali
BioCOLLE...1fields.csv	22 Sep 2014 2:34 pm	2 KB	comm...vali
BioSPECI...31fields.csv	22 Sep 2014 2:34 pm	2 KB	comm...vali
BIOUpload.xls	22 Sep 2014 2:34 pm	6 KB	Micros...ksl
CustomFie...stomer.csv	22 Sep 2014 2:34 pm	49 bytes	comm...vali
CustomFieldData.csv	22 Sep 2014 2:34 pm	75 bytes	comm...vali
CustomFie...llValid.csv	22 Sep 2014 2:34 pm	403 KB	comm...vali

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 998877,9,1/01/2001
```

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 "Next", "Cancel", and "Finish" buttons. 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. It is worth noting the the custom field uploader does not indicate that the data will be updated Vs inserted (this was a design decision made due to the time it would take to analyse all data in validation and again in insertion. But we have an outstanding JIRA issue to re-evaluate this decision <https://the-ark.atlassian.net/browse/ARK-1444>).

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, and Subject Upload. The current page is Step 3/5: Data Validation. A message states: "The data in the file is now validated, correct any errors and try again, otherwise, click Next to continue." Below this is a section titled "Validation Messages:" which contains an empty box. Under "Data Preview: (First 20 rows of file)" is a table:

ROW	SUBJECTUID	CANCER_BREAST	CANCER_DX_DATE
1	998877	9	1/01/2001

At the bottom, there is a legend: a red square indicates error, an orange square indicates data, a green square indicates data, and a blue square indicates subject.

If things look good, 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 displays the message 'Step 4/5: Confirm Upload' and '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. The central message box now displays '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".' The 'Finish' button is visible at the bottom right of the message box.

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.

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 selected * 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:	Status:
538	SubjectCustomFieldUploadForWhiteLifePool.csv	CSV	arksuperuser@ark.org.au	2015-03-26 13:18:13.0	2015-03-26 13:18:13.0	<input type="button" value="Download Subject Template"/> <input type="button" value="Download Custom Field Template"/> <input type="button" value="Download Consent Field Template"/> <input type="button" value="Download Pedigree Template"/>
537	SubjectCustomFieldUploadForWhiteLifePool.csv	CSV	arksuperuser@ark.org.au	2015-03-26 12:51:45.0	2015-03-26 12:51:45.0	<input type="button" value="Download File"/> <input type="button" value="Download Report"/>
536	SubjectCustomFieldUploadForWhiteLifePool.csv	CSV	arksuperuser@ark.org.au	2015-03-26 12:15:18.0	2015-03-26 11:51:27.0	<input type="button" value="Download File"/>
535	SubjectCustomFieldUploadForWhiteLifePool.xls	XLS	arksuperuser@ark.org.au	2015-03-26 11:50:31.0	2015-03-26 2015-03-24	<input type="button" value="Download File"/>
534	SubjectCustomFieldUploadForWhiteLifePool.csv	CSV	arksuperuser@ark.org.au	2015-03-26 2015-03-24	2015-03-26 2015-03-24	<input type="button" value="Download File"/>

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 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.

Welcome arksuperuser@ark.org.au | Logout

Study: [Study ID] uploadReport537.txt uploadReport538.txt

Study: the-ark.org.au Upload Report
Thu Mar 26 13:18:12 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
0 fields were inserted.
2 fields were updated.

Subject Upload

Choose One *

g the data prior to uploading.

sh

Download Options:

- Download Subject Template
- Download Custom Field Template
- Download Consent Field Template
- Download Pedigree Template

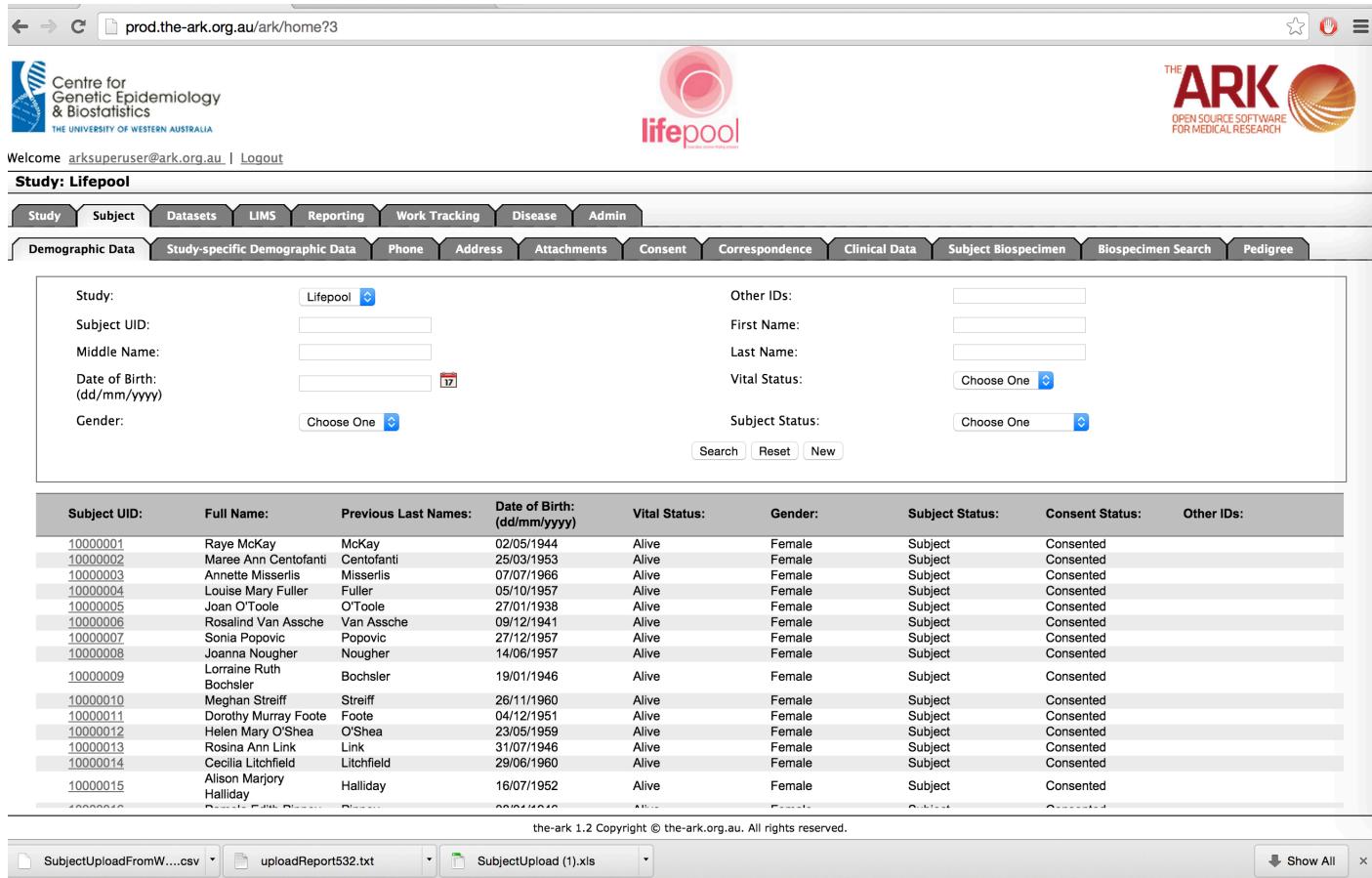
Time: 03-26 13:18:13.0 13:18:13.0

Status: Successfully completed

ID	File Name	Type	User	Upload Time	Process 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

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

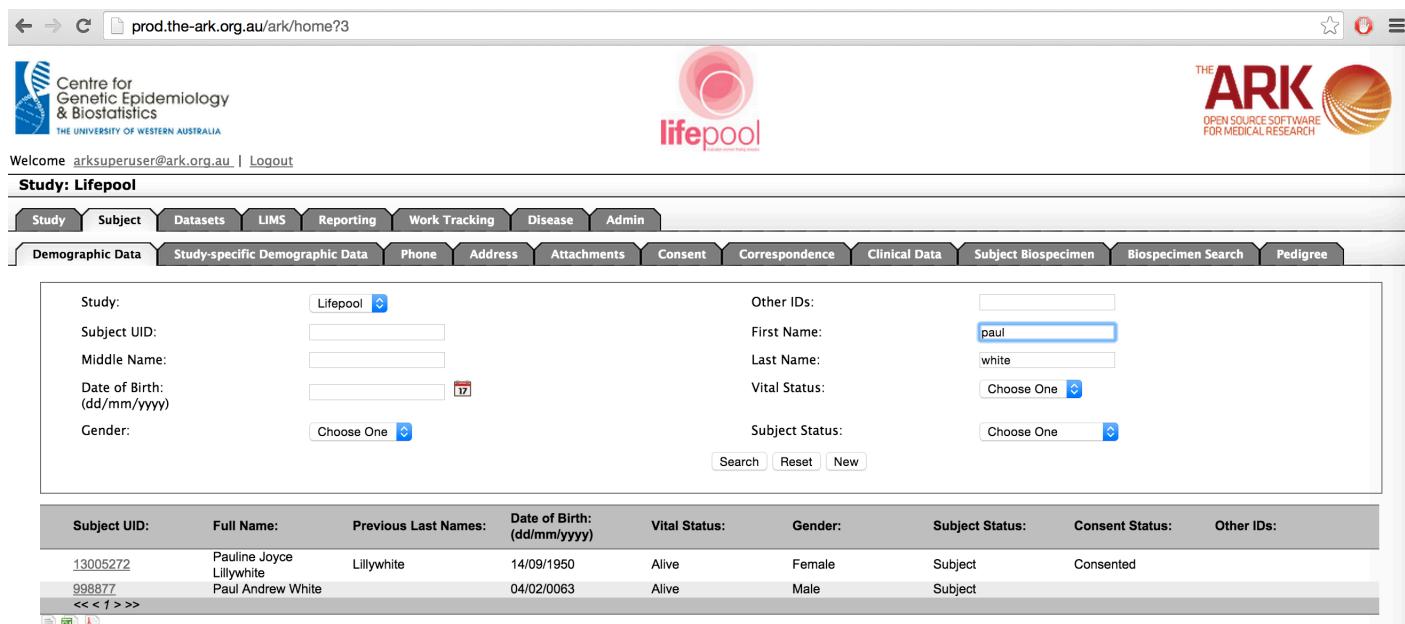
Study: Lifepool

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	14/09/1950	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 three logos: 'Centre for Genetic Epidemiology & Biostatistics' (University of Western Australia), 'lifepool' (with a red circular logo), and 'THE ARK OPEN SOURCE SOFTWARE FOR MEDICAL RESEARCH' (with a stylized orange globe logo). Below the logos, the welcome message 'Welcome arksuperuser@ark.org.au | Logout' is displayed. The main header indicates the study is 'Lifepool' and the subject's UID is '998877'. A navigation bar below the header includes links for 'Study', 'Subject', 'Datasets', 'LIMS', 'Reporting', 'Work Tracking', 'Disease', and 'Admin'. Under the 'Subject' link, a sub-menu is open with options: 'Demographic Data' (selected), 'Study-specific Demographic Data', 'Phone', 'Address', 'Attachments', 'Consent', 'Correspondence', 'Clinical Data', and 'Subject Biospecimen'. The 'Demographic Data' section contains various input fields: 'Send subject a newsletter?' with a dropdown menu 'Choose One'; 'Medicare number' (empty field); 'Medicare dataset consent' with a dropdown menu 'Choose One'; 'Subject diagnosed with breast cancer?' with a dropdown menu 'Unknown'; 'Breast cancer diagnosis date' showing '01/01/2001' with a calendar icon; 'Source of breast cancer diagnosis' with a dropdown menu 'Choose One'; 'Other source of breast cancer diagnosis' (empty field); 'Surgeon providing treatment of breast cancer' (empty field); 'Clinic of the surgeon providing treatment' (empty field); and 'Address of the surgical clinic' (empty field).

We can see that our data has been updated correctly.