Understanding Society User Support - Support #2295

Merging datasets

11/17/2025 01:56 PM - Yashi Jain

Status:	Feedback	**Start date:**	11/17/2025
Priority:	Urgent	**% Done:**	80%
Assignee:	Understanding Society User Support Team		
Category:	Data management		

Description

I am facing issue while merging the data. I have created one dataset using well-being indicators from indresp files for all waves. This dataset also has hidp, msoa and lad. Second dataset is for cultural indicators for only two waves "b" and "e" and I have kept them separately as two years.

Now, when I am merging wellbeing all wave with culture wave b using person_id which is pidp only 67 obs are matching. Please help me what is causing the issue in merge. Do I have to use some other variable to merge?

merge m:1 person_id using "C:\WISERD\Culture\Culture_Ind_KEY_1011.dta"
(variable person_id was float, now double to accommodate using data's values)

```
Result                      Number of obs
    -----------------------------------------
    Not matched                       577,701
        from master                   527,539  (_merge==1)
        from using                     50,162  (_merge==2)

Matched                                   67  (_merge==3)
    -----------------------------------------
for cultural wave e:
```

. merge m:1 person_id using "C:\WISERD\Culture\Culture_Ind_KEY_1314.dta"
(variable person_id was float, now double to accommodate using data's values)
(variable household_id was float, now double to accommodate using data's values)

```
Result                      Number of obs
    -----------------------------------------
    Not matched                       566,802
        from master                   527,115  (_merge==1)
        from using                     39,687  (_merge==2)

Matched                               491  (_merge==3)
    -----------------------------------------
```

History

#1 - 11/21/2025 12:35 PM - Understanding Society User Support Team

- Category set to Data management

- Status changed from New to Feedback

- % Done changed from 0 to 80

- Private changed from Yes to No

Dear Yashi,

It's hard to know for sure without seeing the full data management process that produced these datasets. My best guess is this: "(variable person_id was float, now double to accommodate using data's values)". This suggests that you converted pidp, which is originally a long (because it can contain very large numbers), into a float. If I'm not mistaken, a float can accurately store only around seven digits – and many pidps are longer than that.

In other words, the identifier variable was likely corrupted somewhere along the way.

To check this, try linking to xwavedat using your personal identifier. That file contains all respondents ever enumerated or interviewed, so if you cannot find matches for all individuals in your dataset, the identifier has definitely been damaged.

I hope this helps.

Best wishes,
Piotr Marzec
UKHLS User Support Team