

## Session 4: Export

At the end of this session participants will be able to:

- Use the export data tool to take data from CSPro to other software packages
- Create relations in the dictionary to join records for export

### The Export Data Tool

The CSPro export data tool is available from the Tools menu. When you first start Export Data you are prompted to provide a data dictionary. Choose the PopstanLFS dictionary. From the main export screen, you can choose which records/variables to export using the checkboxes next to each one. To start with, let's just pick the id items and the first few fields from household and geographical identifiers. At the bottom of the export window you can choose the file format to export to. For this exercise, we will choose CSV which can be easily opened in Excel. To run the export, click on the traffic light. You are prompted to choose the data file. We will use the PopstanLFS.csdb data file. Finally, you are prompted for the name(s) of the exported files. Once the export completes the exported files are displayed in Text Viewer. Let's open them in Excel and see what we have. Note that each household is saved in a separate line in the Excel file.

PROVINCE	DISTRICT	GHDWNO	GHHHNO	GHAREA	GHVFDT	GHRPL
1	1	1	1	1	20170918	2
1	1	1	2	1	20170918	2
1	1	1	3	1	20170919	2
1	1	1	4	1	20170918	1
1	1	1	5	1	20170908	1

### Exporting Repeating Records

Let's try exporting a few fields from the demographics roster: line number, name and sex. Notice for each variable we exported we get 20 columns. Why is that? We are getting one column for each record occurrence. CSPro is still exporting each household on a single row and since there are up to 20 people in the household, it generates 20 columns for each variable. Even empty occurrences are still generating columns in the spreadsheet.

PPNO_01	PPNO_02	PPNO_03	PPNO_04	PPNO_05	PPNO_06	PPNO_07	PPNO_08
1	2						
1							
1	2	3					
1							
1	3						

1 row = 1 household

Having a column for each occurrence can complicate working with the data. For example, in this file it is rather tough to do something as simple as count the total number of people. If instead of choosing the default setting of putting multiple record occurrences in a single row (the *All in One Record* setting) try selecting *As Separate Records*. Now we get each person in the household in a separate row. It is important to include the household id items when doing this so that it is clear which people are in which household.

PROVINCE	DISTRICT	GHDWNO	GHHHNO	PPNO	NAME	DEMSEX
1	1	1	1	1	John	1
1	1	1	1	1	2Sam	2
1	1	1	1	2	1Marvin	1
1	1	1	1	3	1Melania	2
1	1	1	1	3	2Donald	1
1	1	1	1	3	3Baron	1
1	1	1	1	4	1Laura	2
1	1	1	1	5	1Homer	1
1	1	1	1	5	3Marge	2

## Exporting Multiple Record Types

Now let's try exporting the first few items from both the demographics record and the labor record. With our current settings, Export Data warns us that only items from the first record will be exported. Why? The problem is that if you put each record in its own row then some rows would have demographic records on them and others would have labor records on them but then the columns would not match.

One solution is to export each record in a separate file by selecting Multiple Files (*one for each record type*) under *Number of Files Created*. Doing this generates two files: DEM\_REC.csv which contains the demographic records and LABOR\_REC.csv which contains the labor records.

DEM\_REC.csv

PROVINCE	DISTRICT	GHDWNO	GHHHNO	PPNO	NAME	DEMSEX
1	1	1	1	1	John	1
1	1	1	1	2	Sam	2
1	1	1	2	1	Marvin	1
1	1	1	3	1	Melania	2
1	1	1	3	2	Donald	1
1	1	1	3	3	Baron	1
1	1	1	4	1	Laura	2
1	1	1	5	1	Homer	1
1	1	1	5	3	Marge	2

LABOR\_REC.csv

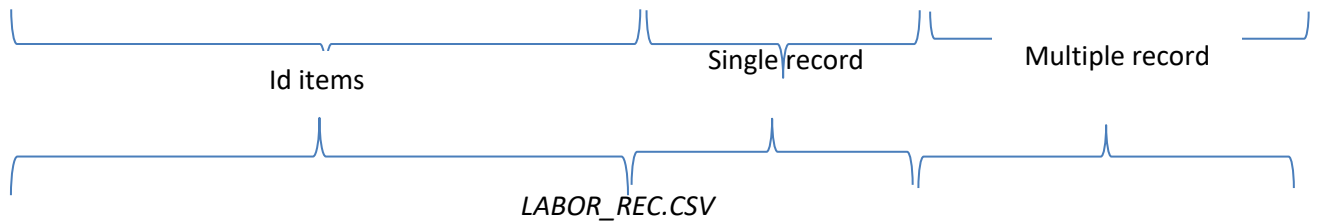
PROVINCE	DISTRICT	GHDWNO	GHHHNO	LBPPNO	PXYRSP	PXYCHK	PXYPPN
1	1	1	1	1	1	1	
1	1	1	1	2	2	2	1
1	1	1	2	1	1	1	
1	1	1	3	1	1	1	
1	1	1	3	2	1	1	
1	1	1	4	1	1	1	
1	1	1	5	1	1	1	
1	1	1	5	2	1	1	

The individuals in these two files can be linked together based on the id items.

It also possible to have Export Data join together single and multiple records. If you select this option, then one file will be generated for each repeating record type and the selected columns for all the selected singly occurring records will be added to each row in each of the exported files. For example, if we choose GHAREA and GHVFDT from the single geographic and identifiers records and also choose the first few items from the demographics record then GHAREA and GHVFDT will be added to each of the two exported files: DEM\_REC.csv and LABOR\_REC.csv.

DEM\_REC.CSV

PROVINCE	DISTRICT	GHDWNO	GHHHNO	GHAREA	GHVFDT	PPNO	NAME	DEMSEX
1	1	1	1	1	20170918	1	John	1
1	1	1	1	1	20170918	2	Sam	2
1	1	1	2	1	20170918	1	Marvin	1
1	1	1	3	1	20170919	1	Melania	2
1	1	1	3	1	20170919	2	Donald	1
1	1	1	3	1	20170919	3	Baron	1
1	1	1	4	1	20170918	1	Laura	2
1	1	1	5	1	20170908	1	Homer	1
1	1	1	5	1	20170908	3	Marge	2



LABOR\_REC.CSV

PROVINCE	DISTRICT	GHDWNO	GHHHNO	GHAREA	GHVFDT	PXYRSP	PXYCHK	PXYPPN
1	1	1	1	1	20170918	1	1	
1	1	1	1	1	20170918	2	2	1
1	1	1	2	1	20170918	1	1	
1	1	1	3	1	20170919	1	1	
1	1	1	3	1	20170919	1	1	
1	1	1	4	1	20170918	1	1	
1	1	1	5	1	20170908	1	1	
1	1	1	5	1	20170908	1	1	

## Using Relations to Join Records

Normally the export data tool will not allow you to export both the demographic and labor records into the same file. In order to do that we must modify the dictionary to add a relation that links the two tables. To create a relation, in the dictionary editor choose “Relations” from the edit menu and then click the add button. For the name enter DEM\_LABOR\_REL. For the primary record, choose DEM\_REC linked by PPNO and for the secondary record choose LABOR\_REC linked by LBPPNO. Now close and reopen the export data tool to get the updated version of the dictionary. You will now see the relationship shown in export data. You can now choose variables from inside the relation that come from both of the linked records and they will be exported into the same file.

DEM\_LABOR\_REL.CSV

PROVIN CE	DISTRI CT	GHDW NO	GHHHN O	PPN O	NAME	DEMS EX	LBPPN O	PXYR SP	PXYC HK	PXYPP N
1	1	1	1	1	John	1	1	1	1	
1	1	1	1	2	Sam	2	2	2	2	1
1	1	1	2	1	Marvi n	1	1	1	1	
1	1	1	3	1	Melan ia	2	1	1	1	
1	1	1	3	2	Donal d	1	2	1	1	
1	1	1	3	3	Baron	1				
1	1	1	4	1	Laura	2	1	1	1	
1	1	1	5	1	Home r	1	1	1	1	
1	1	1	5	3	Marge	2				

In some cases it may be easier to do the export of the records without joining and then do the join in the software that you have imported the data into.

### Importing Data into SAS, SPSS, Stata and R

When exporting data to statistical packages, CSPro generates both a data file and script to run inside the statistical software itself to run the import. For details in how to run this script for each package see the online help for Export Data and look under “How to...”.

### Saving your Export Specification

You can save your export settings as CSPro export specification file (.exf). You can later double click on this file or open it from the Export Data Tool to retrieve all the selections that were made.

## Exercises

1. Using the test data, export the demographics record along with the ID items to the package of your choice (Excel, SPSS, Stata, SAS or R). Use that package to determine the total number of people by sex (total, males, females) and the average household size.
2. Export all three records into a single file. Use the result to determine the number of males who responded directly to the questions themselves and were in households interviewed on September 18.