

COMPSOC

*exploiting, documenting, and enriching
COMParative data from large-scale surveys in the
SOCial sciences*

CESSDA PPP Expert Workshop on Harmonisation
Paris 3 April 2008

Ruud Luijkx
Tilburg University

R.Luijkx@uvt.nl

Background

- Funded by the Dutch National Science Foundation
- Partners are:
 - Tilburg University (Jacques Hagedaars, Ruud Luijkx, Bas Kurver, Ellen Verbakel)
 - Twente University (Kees Aarts, Jacques Thomassen, Bojan Todojievic a.o.)
 - Radboud University Nijmegen (Rob Eisinga, Nan Dirk de Graaf, Stijn Ruiter a.o.)
- In collaboration with DANS
- Until June 2009

Surveys in the project

- European Values Study (EVS)
 - 3 waves (1981, 1990, 1999), repeated cross-sections
 - topic: norms and values (religion, work and leisure, marriage and family, politics)
- Social Cultural developments in the Netherlands (SOCON)
 - 6 waves (1979-2005), repeated cross-sections
 - topic: religiosity, politics and immigration
- Dutch Election Studies (NKO)
 - 11 waves (1971-2006), repeated cross-sections (some panel)
 - topic: electoral behaviour, party identification
- Family survey Dutch Population (FNB)
 - 4 waves (1992-2003), repeated cross-sections
 - topic: life course and life situation

COMPSOC aims

1. Making the four data sets better accessible in their longitudinal form (in NESSTAR) with extensive meta information up to the higher standards
2. Enriching the data files:
 - “interconnecting” the data sets;
 - add aggregate characteristics (national, historical, birth cohort) from (1) the other four data sets (2) other micro data (3) “administrative” sources
3. To publish (methodological) articles on using the enriched data.

1. Publication of all data

- Micro data in its longitudinal form (SPSS, STATA, SAS)
- Meta data (DDI 1.2.2)
- Publication in NESSTAR
 - Highly accessible
 - Clear overview of data (groupings)
 - Downloadable (selections of) micro data
 - Availability of metadata
 - But not very well designed for longitudinal/comparative data

2. Enriching data files

- Information from other micro data:
 - aggregates from the “ own” social surveys, e.g. opinions by age/birth cohort (by sex, etc.)
 - aggregates are means or standard deviations
- Information from administrative sources
 - national and regional statistics (over time)
 - e.g. % immigrants; % unemployment

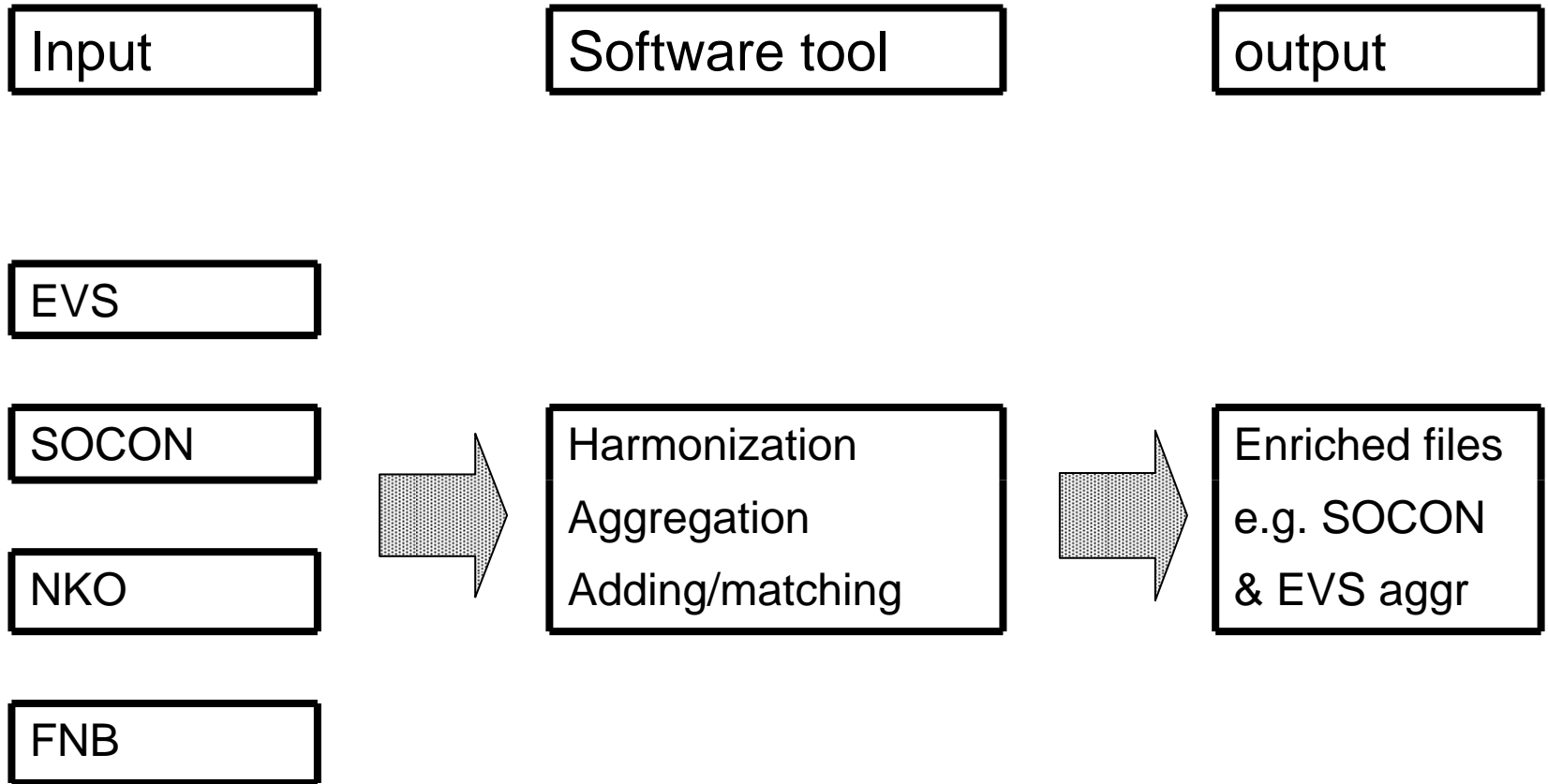
Harmonisation needed

- Within surveys
 - Standardize items between years, but retain original variables as well
- Between surveys
 - Standardize items among (longitudinal) surveys, but retain original variable as well
 - THIS WILL BE AN INFINITE TASK when taking many surveys into account; therefore we will need some SOFTWARE TOOL here.

Software tool (1)

- Utility that can:
 - Harmonize variables between micro data files
 - Aggregate micro data
 - Add aggregates to micro data files (add & match)
 - Ideally records all changes in DDI output
- Resulting in:
 - A workable customized micro data file
 - Documentation (DDI) of all variables including the harmonization process (ideally from DDI 1.2.2 to DDI 3.0)

Software tool (2)



Software tool (3)

Input

Software tool

Output

EVS	
Age	Abortion
18	1
20	2
40	4
45	5

SOCON	
Age	Left/right
19	1
21	4
41	5
45	7

Software tool	
Age grouping	Aggregate
(10 yrs)	(mean)
16 thru 25 eq 1	$(1+2)/2=1,5$
36 thru 45 eq 2	$(4+5)/2=4,5$
Age grouping	
(10 yrs)	
16 thru 25 eq 1	-
36 thru 45 eq 2	-

SOCON & EVS aggr		
age	Left/right	Mean abortion
19	1	1,5
21	4	1,5
41	5	4,5
45	7	4,5

Software tool (4)

