**Description of the SDR2 Missing Values Coding Schema**

In the SDR2 MASTER file we identify different situations that warrant to be treated as missing data. We propose a schema of coding missing values (see Table 1) that is a trade-off between (1) what we would like to see in other survey projects and (2) what was feasible in SDR2, given the variety of missing data situations in the source data files. Moreover, (3) target categories should also be straightforward to interpret by both SDR coders and end-users.

*Missing data* in the MASTER file are of two types. The first type contains all missing data *declared in the source surveys*. Using data providers' descriptions, in the SDR project we distinguish, on the individual level, the following categories: “Don't know” (DK), “No answer” (NA), “Refusal” (REF), “Don't understand the question” (DU), any combination of DK, NA, REF, DU (denoted by DNR), and “Inapplicable” (INAP). Labels and specific values for all these situations differ among source surveys. The SDR harmonization process standardizes the variety of codes for missing values, providing a uniform schema (see Table 1, Harmonized source codes for missing values).

The second type of missing data stems from *ex-post harmonization.* We created new missing value codes whenever we did not find, for a given source survey, suitable item(s) for the target variable (“Question not asked” or QNA), or the source value did not fit the target (UNFIT), or the whole set of an item’s response options were insufficiently defined (INSUF). Sometimes the surveys contained undocumented missing values or obvious errors in the data (ERR) or the provided description did not allow us to assign it to any other category than “Not elsewhere classified” (NEC). COMBI catches a very specific situation, when we use two source variables for creating a target and on both variables a respondent does not give a valid answer, however the source for missing is different, say “Don’t Know” on the first item and “Refused” on the second. Finally, we use the CINAP code only for harmonization control variables, when the target is missing (for any reason) and the control has no meaning.

Table 1 presents the SDR2 missing values coding schema. Since we share the MASTER file in both STATA and SPSS formats and these programs differ in how they present missing values, our schema includes two sets of missing value codes: dot-letters (.a, .b, etc.), provided in the STATA MASTER file, and negative integers (-1, -2, etc., so that -1 corresponds to .a, -2 to .b, etc.) in the SPSS MASTER file.

In exceptional situations, when we could not use any codes for missing values, we applied an <empty> string or <null> value, depending on the variable's type. This pertains to the following variables:

- T_PROJECT_WAVE (text data type, <empty> strings possible) because some survey projects are standalone and therefore require no wave identification;
- QF_AGE_BIRTHYR_DIFF (integer data type, <null> values possible) has valid values ranging from -80 to +81. Thus, we cannot use negative values as missing values in SPSS.

Note that the <null> value in STATA is represented by the system missing dot value (.)

Technical variables other than T_PROJECT_WAVE (i.e. T_DATAFILE, T_ID, T_PROJECT_NAME, T_COUNTRY_L1U, T_COUNTRY_L2U, T_INT_YEAR, T_COUNTRY_YEAR, T_WEIGHT_L2U) as well as quality control variables other than QF_AGE_BIRTHYR_DIFF (i.e. QR_AGE_SUSPECT, QR_BIRTHYR_SUSPECT, QR_HH_SIZE_OUTLIER, QR_DUPLICATE, QR_ID_NONUNIQUE, QR_ID_MISSING) do not need missing values since they always take valid data.

Note that missing codes schema does not apply to PLUG-COUNTRY and PLUG-SURVEY files.

In the Appendix we provide examples of source missing data labels under specific SDR2 missing data codes.

Table 1. Codes for missing values in SDR2

| SDR2 code [a] | SPSS (STATA) value | Label |
|---|---|---|
| | | **Harmonized source codes for missing values** |
| DK | -1 (.a) | Don't know |
| NA | -2 (.b) | No answer |
| REF | -3 (.c) | Refusal |
| DU | -4 (.d) | Don't understand the question |
| DNR | -5 (.e) | Any combination of DK, NA, REF, DU |
| INAP | -6 (.f) | Inapplicable |
| NEC | -7 (.g) | Not elsewhere classified |
| | | **SDR created codes for missing values** |
| UNFIT | -8 (.h) | Source value does not fit to target |
| ERR | -9 (.i) | Errors in source data and undocumented source values |
| COMBI | -10 (.j) | Different missing codes on multiple sources taken for a target |
| CINAP | -11 (.k) | For control variables only: inapplicable |
| INSUF | -12 (.l) | For survey: Insufficiently defined response categories |
| QNA | -20 (.t) | For survey: Question not available |

[a] Abbreviations for the labels corresponding to the SDR2 missing values. These codes are used in the Cross-walk Tables (CWT files) that accompany documentation of SDR2 target variables.

Blue = clearly defined response on individual level
Green = other response on individual level
Brown = survey / response-options level (CINAP = it depends on item)

Appendix. Examples of source response options assigned to specific SDR2 missing codes

DK
==========
both
can't choose
can't determine
can't remember
can't say
difficult to say
dk
dk-too early to say
don't know
don't know (do not read)
don't know / not applicable
don't know, uncertain
don't know/can't remember
don't know/haven't heard enough
don't remember
don't remember what i did/don't know
forget
hard to say
hasn't heard enough / dk
haven't thought much about it
i don’t remember what i did
it depends
it depends (spontaneous)
mentioned // e.g. EB_35 item: v245
n/s
no opinion
no sabe
not know
nsp, wn
respondent could not recall

NA
==========
+ // CNEP_3_ES item: Z.Sp.F.AsoNA
<na> // if explained in documentation
da
don't answer
dr
ka // a typo in EB_9 item: v70
n.c.
n/r
na // if explained in documentation
no answer

no answer (es, it: incl. dk or don't remember)
no answer, nl: no answer, refused
no answer/<na>
no answer; nl: na, above 30 years at school
no reply
no response
not answered
not applicable / no answer
not specified
not stated
nr
unspecific
unspecific answer

REF
==========
da
decline to answer
decline to answer (do not read)
no more answer
ref
refusal
refusal (spontaneous)
refusal / not applicable
refusal / null
refuse to say
refuse to state
refused
refused to answer
yes // e.g. CNEP_4_US item: Q17_Refused

DU
==========
do not recognize
do not understand question / what 'democracy' means
do not understand the question
not clear

INAP
==========
break off
filtered
inap
inapplicable
n/a // if explained in documentation
nap
nap,nav

nav
nav/na
not applicable
not available
not available (tw:not asked for subsample)
such an association does not exist here

NEC
==========
<na> // if not explained in documentation
<na> <not coded>
age not registered
gender not registered
informal schooling only // education level items in AFB
interviewer error
literate but no formal schooling // CNEP_3_TW item: L.Education
missing
missing data
missing value
missing; unknown
n.a. // if not explained in documentation
na // if not explained in documentation
na (dk/na - lt, w2)
na, others
no answer, information missing
no answer, other
no answer, other qualification, education
no answer/no data
no data
no registration
not ascertained
not clear
not concerned / not interested
not important
not known
not registered
not relevant (spontaneous)
other
other answer
other educ,other answer ch: other education
other responses
other responses, dk, etc.
refused, rural // ISSP_1986 item: v82

UNFIT
==========
// civil status specific items

engaged
have a partner / living as married
legally divorced/civil union dissolved / none of these (never married or in legally registered civil union)
legally married / none of these (never married or in legally registered civil union)
no partner / living as married
no partner / married

// education specific items
completed education abroad
foreign and other (overseas school leaving exams)
other (including other schools)
other education
other qualifications
still at school, no formal education // in ISSP_2004 item: v204 (only for Finland)

// household-size specific items
2 people lived in “a collective accommodation for aged persons“
none member
not a private household

// income specific items
can't calculate
he/she is single
inap/no income
no income, d.k., n.a.
no income; not available
no information on family income
unreliable response
value could not be calculated - only applies in 2011

// rurality / settlement-size specific items
hostel
other country
refugee camp
smallholdings

// voting specific items
vote blank,not eligible