SDR2 PLUG_SURVEY File Overview

1. General Information

The auxiliary survey-level dataset, PLUG-SURVEY is part of the SDR2 database (see SDR2 Database Overview) and provides three types of source survey indicators – (1) quality controls for source survey documentation, (2) data records, and (3) processing errors (cf. Table 1 and Section 2) – next to a set of technical variables needed for merging this file with the MASTER file. Source quality indicators for survey documentation and data records capture information at the national survey level and for processing errors - at the survey wave level. One exception is the processing error index, QP_PER_INDEX, which measures the rate of processing errors at the project-wave level (see Section 2.3).

In the MASTER file, national surveys are uniquely identified by the following technical variables: T_PROJECT_NAME, T_PROJECT_WAVE, T_COUNTRY_L1U, T_COUNTRY_L1U, and T_COUNTRY_YEAR. For details about the MASTER file, technical variables and survey selection, see the SDR2 MASTER File Overview.

Variables in the PLUG-SURVEY file are provided for all surveys selected in the MASTER file.

Table 1. Description of variables in the SDR2 PLUG-SURVEY file

Variable name	Variable description	Variable values [a]
Technical variables		
T_PROJECT_NAME	Survey project name	
T_PROJECT_WAVE	Survey project wave	
T_COUNTRY_L1U	Code for country/territory unit, level 1	
T_COUNTRY_L2U	Code for country/territory unit, level 2	
T_COUNTRY_YEAR	Year in which national survey was conducted	
Source survey quality: survey documentation accuracy		
F_SAMPLECOVER_SVY	Flag: Population coverage in source sample	1 = Sample drawn from entire country 2 = Sample drawn from part of country (samples for all parts available) 3 = Sample drawn from part of country (samples for some parts not available) 4 = Country sample extracted from larger territory 5 = Sample not representative for country (based on documentation)
QD_UNIVINFO_SVY	Documentation provides information about target population (universe)	1 = Yes 0 = No
QD_SAMPLEINFO_SVY	Documentation provides information about sampling procedure	0 = No information on sample type 1 = Insufficient information to identify sample type 2 = Sample type identified, description limited 3 = Sample type identified, description detailed

QD_SAMPLETYPE_SVY	Type of sampling schema, cf. documentation [nominal variable, see Section 2.1 and Table 2]	0 = No documentation about the sample at all … 52 = Non-quota non-representative samples
QD_RRVALUE_SVY	Response rate (RR)	[decimal value]
QD_RRINFO_SVY	Documentation provides information about RR	1 = Yes 0 = No
F_RRAPPROX_SVY	Flag that documentation provides approximated RR value	0 = RR value not approximated 1 = RR value is approximated 2 = More than one RR value provided
F_RRCALC_SVY	Flag that RR value is derived by the SDR team	1 = Yes 0 = No
QD_TRANSLINFO_SVY	Documentation provides information about translation method	1 = Yes 0 = No
QD_TRANSLVALUE_SVY	Documentation specifies use of professional translation method	1 = Yes 0 = No
QD_PRETINFO_SVY	Documentation discusses pretesting or piloting	1 = Yes 0 = No
QD_PRETVALUE_SVY	Documentation specifies that pretesting was performed	1 = Yes 0 = No
QD_FCTRLINFO_SVY	Documentation discusses fieldwork control	1 = Yes 0 = No
QD_FCTRLVALUE_SVY	Documentation specifies that fieldwork control was performed	1 = Yes 0 = No
QD_MODEINFO_SVY	Documentation provides information about interview mode	1 = Yes 0 = No
QD_DOCU_INDEX	Summary index, source survey documentation quality	0 = Lowest quality 7 = Highest quality
Source survey quality: errors in data records in source data files		
QR_DUPLICATE_SVY	Survey contains non-unique records (NUR)	1= True 0 = False
QR_ID_ERROR_SVY	Survey contains errors in case IDs	1= True 0 = False
QR_HIMISS_AGE_GENDER_SVY	Survey has over 5% missing data on source age or gender	1= True 0 = False
QR_WGHTTYPE_SVY	Type of source weight	0 = No weight 1 = Design weight 2 = Post-stratification weight 3 = Combined (design and post-stratification) weight 4 = No information 5 = Constant weight
QR_WGHTDESC_SVY	Status of source weight variable	0 = Source variable not available, target value set to 1 3 = Source variable not available, target value set to a proportional coefficient

		10 = Source value is treated as the target 15 = Source value used after rescaling for sample size 20 = Sample shares a single source value (1) used as the target 25 = Sample shares a single source value (1) used as the target after rescaling 30 = Sample shares a single value, but using a proportional coefficient 35 = Sample shares a single value, but using a proportional coefficient, rescaled 40 = Multiple source values transformed by a proportional coefficient 45 = Multiple source values transformed by a proportional coefficient, rescaled
QR_WGHTDISTORT_SVY	Sample size change when source weights applied	[numerical value]
QR_RECORDS_INDEX	Summary index, data records error types	0 = No errors (highest quality) 4 = Largest number of error types (lowest quality)
Source survey quality: processing errors		
QP_GENDER	Quality control processing error types: Gender (realized values)	0 = No processing errors 3 = Contradictory variable values 5 = Lack of variable value labels
QP_AGE	Quality control processing error types: Age (realized values)	0 = No processing errors 1 = Illegitimate variable values 2 = Misleading variable values 3 = Contradictory variable values 5 = Lack of variable value labels 7 = Illegitimate & contradictory variable values 9 = Illegitimate & lack of variable value labels 10 = Contradictory variable values & lack of variable value labels 11 = Discrepant & lack of variable value labels 12 = Illegitimate & contradictory & lack of variable value labels
QP_BIRTHYR	Quality control processing error types: Birth year (realized values)	0 = No processing errors 1 = Illegitimate variable values 2 = Misleading variable values 5 = Lack of variable value labels 6 = Illegitimate & misleading variable values 7 = Illegitimate & contradictory variable values 9 = Illegitimate & lack of variable value labels 10 = Contradictory variable values & lack of variable value labels

QP_EDU	Quality control processing error types: Education level (realized values)	0 = No processing errors 2 = Misleading variable values 3 = Contradictory variable values 5 = Lack of variable value labels 10 = Contradictory variable values & lack of variable value labels
QP_EDU_YEARS	Quality control processing error types: Education schooling years (realized values)	0 = No processing errors 1 = Illegitimate variable values 3 = Contradictory variable values 4 = Discrepant variable values 5 = Lack of variable value labels 7 = Illegitimate & contradictory variable values 8 = Illegitimate & discrepant variable values 9 = Illegitimate & lack of variable value labels 10 = Contradictory variable values & lack of variable value labels 11 = Discrepant & lack of variable value labels 12 = Illegitimate & contradictory & lack of variable value labels 13 = Illegitimate & discrepant & lack of variable value labels
QP_TRPARL	Quality control processing error types: Trust in parliament (realized values)	0 = No processing errors 3 = Contradictory variable values 4 = Discrepant variable values 5 = Lack of variable value labels 10 = Contradictory variable values & lack of variable value labels 11 = Discrepant & lack of variable value labels
QP_DEMONST	Quality control processing error types: Participation in demonstration (realized values)	0 = No processing errors 3 = Contradictory variable values 5 = Lack of variable value labels 10 = Contradictory variable values & lack of variable value labels
QP_PER_INDEX	Quality control processing error: Index (errors)	0 = No processing errors (highest quality) 1 = Largest number of processing errors (lowest quality)

[a] Missing values are assigned according to the SDR2 missing codes schema, provided in Appendix 1. For the full documentation of SDR2 missing codes schema, see SDR2 Missing Codes Schema document.

2. General rules and procedures for constructing the controls for source data quality

2.1. Quality controls for source survey documentation (see Table 1) capture methodological features of the survey process that generated the data we harmonized in the SDR 2 project. Specifically, we identified information pertaining to universe, sample, response rate, translation, pretesting, fieldwork control and interview mode.

To do so, we analyzed the source survey documentation. By documentation we understand a variety of descriptive documents, including technical reports, methodological reports, fieldwork descriptions, codebooks, and questionnaires. Data producers provide these files via projects' websites, websites of organizations that led the projects, or through data archives (i.e. GESIS, ICPCR, and UK Data Service).

The source survey documentation is available either per country (e.g. WVS) or per wave (e.g. ISSP and ESS). To identify and code information about survey characteristics according to predefined schemas, different SDR2 team members worked independently, and results of their work were checked for intercoder reliability. Problematic cases, such as discrepancies in coding information pertaining to the same indicator by different team members, were discussed to reach adequate solutions.

QD_UNIVINFO_SVY, QD_RRINFO_SVY, QD_TRANSLINFO_SVY, QD_PRETINFO_SVY, QD_FCTRLINFO_SVY, and QD_MODEINFO_SVY are dichotomous variables. They are coded 1 when the documentation explicitly discusses the target population (universe), response rate, translation method, pretesting, fieldwork control, and interview mode, respectively. Note that we assign the value 1 even if the documentation indicates that a given step, for example, fieldwork, was not carried out. These control variables are coded 0 only if the documentation does not contain any mention of these methodological features.

QD_SAMPLEINFO_SVY captures the quality of the information about the sample type and differentiates between no information, insufficient information, and information which is limited or detailed.

QD_SAMPLETYPE_SVY is a nominal variable that records the sampling schema type that a given national survey employed, according to the source survey documentation. It takes values from 0 to 52, cf. Table 2.

The variable QD_SAMPLETYPE_SVY is accompanied by the flag variable F_SAMPLECOVER_SVY that captures population coverage problems in the source samples. For the discussion about the population coverage problems, see the SDR2 MASTER File Overview.

Table 2. QD_SAMPLETYPE_SVY categories

Values	Value labels
0	No documentation about the sample at all
1	Insufficient information about the sample
10	Simple probability samples
21	Stratified probability samples from individuals' registry
22	Area or cluster probability samples from individuals' registry
23	Multi-stage samples from individuals' registry
31	Multi-level samples with determined within-household selection
32	Multi-level samples with undetermined within-household selection
33	Random route samples with determined within-household selection
34	Random route samples with undetermined within-household selection
41	Any multi-level samples with quota respondent selection
42	Random route samples with quota respondent selection
51	Direct quota samples
52	Non-quota non-representative samples

QD_TRANSLVALUE_SVY is a dichotomous variable coded 1 if the source survey documentation indicates that in a given national survey the expert translation was used (i.e. by a translation bureau or professionally trained translators, with the use of the Translation, Review, Adjudication, Pretesting and Documentation (TRAPD) methods, or using back translation). The variable takes the value 0 if the documentation mentions that translation was carried out by research team members, or if no information about translation method is available.

QD_PRETVALUE_SVY and QD_FCTRLVALUE_SVY are dichotomous variables coded 1 if the source survey documentation indicates that pretesting (by any methods, e.g. pilot interviews with a limited number of respondents, cognitive interviews, focus groups, or expert panels) and, respectively, fieldwork control (e.g. via interviews supervision or back-checking), were carried out. The value 0 means that the source documentation does not indicate that these procedures occurred.

QD_DOCU_INDEX is a summary measure of the quality indicators of data as reflected in source survey documentation at the level of national survey. The index ranges from 0 (lowest quality) to 7 (highest quality). This measure is constructed as the sum of QD_UNIVINFO_SVY, QD_SAMPLETYPE_SVY (where values 10 and greater are coded 1, otherwise 0), QD_TRANSLVALUE_SVY, QD_PRETVALUE_SVY, QD_FCTRLVALUE_SVY QD_MODEINFO_SVY, and QD_RRINFO_SVY.

QD_RRVALUE_SVY provides response rate (RR) values for national surveys, using information from the source survey documentation (i.e. the documentation provides the exact RR value, or it provides other data, such as intended and achieved sample sizes, number of ineligibles etc., that allow us to calculate the RR value). QD_RRVALUE_SVY can range from 0 to 1, with higher numbers corresponding to higher response rates.

QD_RRVALUE_SVY is accompanied by two flag variables:

- F_RRAPPROX_SVY is a nominal variable ranging from 0 to 2 that indicates if the documentation provides the exact RR value (coded 0), the approximated RR value (coded 1), or more than one RR value (coded 2). In the last case, we chose one single value (cf. rules discussed in Section 3).
- F_RRCALC_SVY is a dichotomous variable. It flags situations (coded 1) when values on QD_RRVALUE_SVY are calculated by the SDR team using information from the source survey documentation. As much as possible, we calculated RR values using the AAPOR RR1 (minimum RR) definition (aapor.org/Education-Resources/For-Researchers/Poll-Survey-FAQ/Response-Rates-An-Overview.aspx).

Response Rates: Special Cases

- In some instances, documentation sources provide two response rate (RR) values for a given national survey. Furthermore, sometimes the documentation provides one exact RR value, together with additional data that allow calculating a RR value (e.g. number of complete and incomplete interviews, non-contacts etc.). We mark these cases by assigning the code 2 to the flag variable F_RRAPPROX_SVY and apply the following rules to select a single RR value on QD_RRVALUE_SVY:
  - if the documentation provides two RR values for the same national survey, we assign the higher value to QD_RRVALUE_SVY;
  - if the documentation provides a single RR value as well as the information that allows to calculate RR, we assign the precise RR value to QD_RRVALUE_SVY;
  - if the documentation provides two different sets of numbers that each allow to calculate RR values, we calculate two RR values and assign the higher value to QD_RRVALUE_SVY. In such case, we code 1 the accompanying flag variable F_RRCALC_SVY (i.e. RR calculated by SDR team).

- For AFB 6 SL, QD_RRVALUE_SVY = 1. This value is based on the following description in the documentation (AFB 6 SL Codebook):

  > "As submitted earlier, SRL has over time had zero refusals in not only Afrobarometer but other surveys. This may not be unconnected with the fact that

almost all respondents think surveys are sponsored by Government. Looking, for instance, at the r6 (and even the r5) dataset, apart from respondents citing Afrobarometer or a research company or an NGO (which they think is the implementing agency for Government), nearly every respondent will say it is Government that sent the fieldworkers to do the job; and I'm sure many citizens normally want to comply with Government in many issues especially when they can."

- For a set of NBB samples that provide different RRs for different ethnic groups within country (rather than a single RR value for the survey as a whole), we calculated RR values (and flagged these instances as F_RRCALC_SVY = 1). For example, for NBB 1 LT we calculated the value on QD_RRVALUE_SVY as follows:

  Lithuanians: 86 (Effective sample = 1173, Interviews completed = 1009)

  Russians: 80.1 (Effective sample = 647, Interviews completed = 518)

  Poles: 78.3 (Effective sample = 619, Interviews completed = 485)

  QD_RRVALUE_SVY = (1009 + 518 + 485) / (1173 + 647 + 619) = 0.824

- For a number of country samples, the original completion rate was provided instead of RR. This occurred in the following samples: ISJP 1991 CZ, ISJP 1991 HU, ISJP 1991 JP, ISJP 1991 RU, ISJP 1991 SK, PA1 AT, PA1 GB-GBN, PA2 DE, PA2 NL, PA2 US, PPE7N AT, PPE7N IN, PPE7N JP, PPE7N NG, PPE7N NL.

  According to the documentation, WVS 4 BD has a RR 95-97%. We assigned QD_RRVALUE_SVY = 0.96 for this survey, and flagged it as F_RRAPPROX_SVY = 1.

- Two different documentation sources provided different RR values for WVS 4 MX, 66% and 74%, respectively. We assigned the higher value (0.74) for this survey on QD_RRVALUE_SVY.

2.2. Quality controls for data records indicate whether a national survey exhibits errors in data records in source data files. Generally, these variables are survey-level aggregates based on quality measures constructed at the respondent level and available in the MASTER file (see the SDR2 MASTER File Overview).

QR_DUPLICATE_SVY is coded 1 if a national survey contains at least one non-unique record (NUR), and 0 if no NURs are present. (In the MASTER file, its corresponding respondent-level measure is s to QR_DUPLICATE).

QR_ID_ERROR_SVY is coded 1 if a national survey contains at least one erroneous respondent ID, and 0 if no errors in the respondent ID were identified. (QR_ID_ERROR_SVY is constructed using information from refers to QR_ID_NONUNIQUE and QR_ID_MISSING in the MASTER file.)

QR_HIMISS_AGE_GENDER_SVY is coded 1 if a national survey has over 5% missing data on the measures of either age or gender, and 0 if this is not the case.

QR_WGHTTYPE_SVY captures the type of the weighting factor, whether it is design (coded 1), post-stratification (coded 2), combined (coded 3), no information on the type is provided (coded 4), a constant value (i.e. shared by all respondents in a survey, coded 5), or there is no weight whatsoever (coded 0).

QR_WGHTDESC_SVY describes the operation (if needed) that the SDR team applied (if needed) on the source weight variable. Generally, the source values can be taken as is, rescaled, or multiplied by a proportional coefficient. For details see T_WEIGHT_L2U_GVR_SDR2 report .

QR_WGHTDISTORT_SVY measures the impact of the weighting factor on a source sample, i.e. whether the sample is artificially increased or decreased. Note that values other than 0 indicate a problem with the source variable calibration.

QR_RECORDS_INDEX is a summary measure of data records error types at the level of a national survey. The index, constructed as the sum of QR_DUPLICATE_SVY, QR_ID_ERROR_SVY, QR_HIMISS_AGE_GENDER_SVY, QR_WGHTDESC_SVY (value 0 is coded 1, otherwise 0), QR_WGHTDISTORT_SVY (any distortion different from 0 is coded 1, otherwise 0), and it ranges from 0 (no errors) to 4 (the largest combination of data records error types).

2.3. Quality control for processing error types.

For seven target variables – *gender*, *age*, *birth year*, *education level*, *schooling years* (socio-demographics), *trust in parliament* and *participation in demonstrations* (substantive target variables) – we constructed quality control indicators pertaining to the incidence of certain types of processing errors in the source variables that we use to construct target variables. We identified these processing errors when comparing information about the source variables available in the survey documentation (codebook, questionnaire) with information about the same source variables in data records in the source data files (data dictionaries).

The resulting data quality indicators are the following nominal variables: QP_GENDER, QP_AGE, QP_BIRTHYR, QP_EDU, QP_EDU_YEARS, QP_TRPARL, and QP_DEMONST.

Each of these methodological indicators can take values from 0, which means no processing errors, up to 13. Values larger than 0 identify what type, or combination of types, of processing errors occur in the source variables associated with a target. We also provide the aggregated measure - the index of processing error types QP_PER_INDEX (see Section 2.3.2).

Generally, in the SDR2 project we check for five types of processing errors that we define as follows:

- **Illegitimate Variable Values** captures values of the source variable that are outside of the range that SDR proposes as acceptable, e.g., respondents' number of education years exceeding 25, and being 6 years old when they completed full-time education.

- **Misleading Variable Values** shows that variable values, as coded in the questionnaires/codebooks, are not congruent with the data records in the source data files.
- **Contradictory Variable Values** captures inconsistencies in how the same variable value is labeled in a codebook vs. a questionnaire or a data dictionary, such as when "don't know" value is recorded as 88 in the codebook vs. 888888 in the data dictionary.
- **Variable Values Discrepancy** concerns situations when several or all values in the same variable are inconsistently labeled in codebook, vs. questionnaire or a data dictionary (i.e., there are reversed scales or contradictory value labels for more than one variable value), e.g., age of responds when finished education is recorded as year ranges in the questionnaire vs. as years in the data dictionary.
- **Lack of Variable Value Labels** concerns undefined "nulls" and codes (variable values) in the source data file that are not explained in any source of documentation defining variables and their values (i.e., "wild codes: for a given variable, such as value 7 in data records while the documentation describes the variable as a 1 to 5 scale).

Correspondingly, the full list of SDR2 codes for Processing Error Types is:

0 = No processing errors
1 = Illegitimate variable values
2 = Misleading variable values
3 = Contradictory variable values
4 = Variable values discrepancies
5 = Lack of variable value labels
6 = Illegitimate & misleading variable values
7 = Illegitimate & contradictory variable values
8 = Illegitimate & discrepancy variable values
9 = Illegitimate & lack of variable value labels
10 = Contradictory variable values & lack of variable value labels
11 = Discrepant & lack of variable value labels
12 = Illegitimate & contradictory & lack of variable value labels
13 = Illegitimate & discrepant & lack of variable value labels

Values 6 to 13 indicate that a combination of different processing error types occurs within one source variable.

All processing error types account for inconsistencies between different metadata elements, such as data records on the one hand (i.e., variable values, variable labels, value labels - in the source data files), and information in the codebook or questionnaires, on the other hand.

Information necessary to perform the consistency check is recorded and analyzed in the variable-specific DVRx files:

T_GENDER_DVR_SDR2.xlxs

T_AGE_BIRTHYR_DVR_SDR2.xlxs

T_EDU_DVR_SDR2.xlxs

T_EDU_YEARS_DVR_SDR2.xlxs

T_TRPARL_DVR_SDR2.xlxs

T_DEMONST_DVR_SDR2.xlxs

DVR files are detailed target variable reports that we prepared in a standardized Excel format. They contain information about source variables pooled from different survey projects, including information for processing error checks. We mostly used English-language master documentation, yet documentation availability for source data files differed. Appendix 2, Table A2.2 shows the availability of documentation sources per data file used for processing error check.

2.3.2. Index of processing error types QP_PER_INDEX

The index of processing error types, QP_PER_INDEX is an aggregate measure of discrepancies between documentation and data records in a source survey wave. The index takes into consideration (a) the results of processing errors evaluation for the source variables used to construct seven specific target variables, i.e. *gender, age, birth year, education level, schooling years, trust in parliament* and *participation in demonstrations*, and (b) how many of these target variables are available in a given source survey wave.

QP_PER_INDEX ranges from 0 to 1, where 0 means there are no processing errors in any of the source variables associated with either of the available target variables in a source survey wave (i.e. highest quality). The value 1 indicates that the source variables corresponding to each available target variable in a source survey wave display at least one type of processing error (i.e. lowest quality).

To calculate the Processing Error Types Index, QP_PER_INDEX, we first recode information from QP_GENDER, QP_AGE, QP_BIRTHYR, QP_EDUCATION, QP_SCHOOLYRS, QP_TRPARL and QP_DEMONST so that 1 means a processing error (of any type) in a survey's source variable or set of source variables used to construct a target, and 0 means that no processing error is identified in a source data file. Then, we compute the index as the total number of processing errors per data file divided by the number of available target variables:

QP_PER_INDEX = SUM (processing errors per data file) / N (target variables per data file)

For those few situations when the source data file does not correspond to the source survey wave, we standardize the QP_PER_INDEX to the survey wave-level.[1] Standardizing the measurement

[1]In SDR2, most source data files (145 out of 215, 67.4%) are wave-level files of a given international survey project (e.g. ESS 1). In six instances – ABS 4, AMB 2016, CNEP 3 and 4, PP7EN and VPCPCE – wave level data files were not available at the time we retrieved the data. Therefore, we exceptionally used national survey level data files. In addition, for six other cross-national projects, namely CDCEE, EQLS, EVS, ISJP, NBB and WVS, we reprocessed information from their updated cumulative data files that include two or more merged waves. For the correspondence of source data files to project waves, see Appendix 2, Table A2.1.

of processing quality indicators to the wave-level entailed (i) aggregating the indicators stemming from source data files containing a single national survey, and (ii) disaggregating the indicators stemming from source files containing the data of multiple project-waves.

Appendix 1: Codes for missing values in SDR2

In the SDR database v.2 we identify different situations that warrant to be treated as missing data. Table A.1 lists all SDR2 missing value codes:

Table A.1. Codes for missing values in SDR2

SDR tag [a]	SPSS (STATA) codes	Label
		Standardized 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 codes for missing values. These tags are used in the Crosswalk Table (CWT) files (Excel) that accompany documentation of SDR2 target variables.

In exceptional situations when codes for missing data listed in Table A.1 cannot be used, we apply a system missing <null> value. For more information, see Schema for Missing Value Codes in SDR2 document.

Appendix 2. Quality control measures of processing error types

Table A2.1. Correspondence of source data files to project waves

Acronym	Survey project name	N of datafiles	N of waves	Time span	Comment
ABS	Asian Barometer Survey	12	4	2001-2016	National level data files for wave 4
AFB	Afrobarometer	6	6	1999-2015	
AMB	AmericasBarometer	35	7	2004-2017	National level data files for wave 2016
ARB	Arab Barometer	4	4	2006-2017	
ASES	Asia Europe Survey	1	1	2000	
CB	Caucasus Barometer	7	7	2008-2015	
CDCEE	Consolidation of Democracy in Central and Eastern Europe	1	2	1990-2001	A single cumulative data file
CNEP	Comparative National Elections Project	14	2	2004-2012	National level data files
EB	Eurobarometer	48	48	1970-2017	
EQLS	European Quality of Life Survey	1	3	2003-2012	A single cumulative data file
ESS	European Social Survey	8	8	2002-2016	
EVS	European Values Study	1	4	1981-2009	A single cumulative data file
ISJP	International Social Justice Project	1	2	1991-1997	A single cumulative data file
ISSP	International Social Survey Programme	31	31	1985-2015	
LB	Latinobarómetro	19	19	1995-2016	
LITS	Life in Transition Survey	3	3	2006-2016	
NBB	New Baltic Barometer	1	6	1993-2004	A single cumulative data file
NEB	New Europe Barometer	7	7	1991-2005	
PA1	Political Action (An Eight Nation Study)	1	1	1973-1976	
PA2	Political Action II	1	1	1979-1981	
PPE7N	Political Participation and	7	1	1966-1971	National level data files

	Equality in Seven Nations				
VPCPCE	Values and Political Change in Postcommunist Europe	5	1	1993	National level data files
WVS	World Values Survey	1	6	1981-2014	A single cumulative data file

Table A2.2. Availability of documentation for 215 source data files used in the processing quality check

N	Source data file	Questionnaire	Codebook	Data dictionary	Comment
1	ABS_1	+	+	+	
2	ABS_2	+	+	+	
3	ABS_3	+	+	+	
4	ABS_4_KH	NA	NA	+	wave level questionnaire
5	ABS_4_KR	NA	NA	+	wave level questionnaire
6	ABS_4_MM	NA	NA	+	wave level questionnaire
7	ABS_4_MN	NA	NA	+	wave level questionnaire
8	ABS_4_MY	NA	NA	+	wave level questionnaire
9	ABS_4_PH	NA	NA	+	wave level questionnaire
10	ABS_4_SG	NA	NA	+	wave level questionnaire
11	ABS_4_TH	NA	NA	+	wave level questionnaire
12	ABS_4_TW	NA	NA	+	wave level questionnaire
13	AFB_1	NA	+	+	
14	AFB_2	NA	+	+	
15	AFB_3	NA	+	+	
16	AFB_4	NA	+	+	
17	AFB_5	NA	+	+	
18	AFB_6	NA	+	+	

19	AMB_2004	+	NA	+	
20	AMB_2006	+	NA	+	
21	AMB_2008	+	NA	+	
22	AMB_2010	+	NA	+	
23	AMB_2012	+	NA	+	
24	AMB_2014	+	NA	+	
25	AMB_2016_AG	+	NA	+	
26	AMB_2016_AR	+	NA	+	
27	AMB_2016_BO	+	NA	+	
28	AMB_2016_BR	+	NA	+	
29	AMB_2016_CA	+	NA	+	
30	AMB_2016_CL	+	NA	+	
31	AMB_2016_CO	+	NA	+	
32	AMB_2016_CR	+	NA	+	
33	AMB_2016_DM	+	NA	+	
34	AMB_2016_DO	+	NA	+	
35	AMB_2016_EC	+	NA	+	
36	AMB_2016_GD	+	NA	+	
37	AMB_2016_GT	+	NA	+	
38	AMB_2016_GY	+	NA	+	
39	AMB_2016_HN	+	NA	+	
40	AMB_2016_HT	+	NA	+	
41	AMB_2016_JM	+	NA	+	
42	AMB_2016_KN	+	NA	+	
43	AMB_2016_LC	+	NA	+	
44	AMB_2016_MX	+	NA	+	
45	AMB_2016_NI	+	NA	+	

46	AMB_2016_PA	+	NA	+	
47	AMB_2016_PE	+	NA	+	
48	AMB_2016_PY	+	NA	+	
49	AMB_2016_SV	+	NA	+	
50	AMB_2016_US	+	NA	+	
51	AMB_2016_UY	+	NA	+	
52	AMB_2016_VC	+	NA	+	
53	AMB_2016_VE	+	NA	+	
54	ARB_1	NA	+	+	
55	ARB_2	NA	+	+	
56	ARB_3	NA	+	+	
57	ARB_4	NA	+	+	
58	ASES	NA	+	+	
59	CB_2008	+	NA	+	
60	CB_2009	+	NA	+	
61	CB_2010	+	NA	+	
62	CB_2011	+	NA	+	
63	CB_2012	+	NA	+	
64	CB_2013	+	NA	+	
65	CB_2015	+	NA	+	
66	CDCEE_1_2	NA	+	+	
67	CNEP_3_ES	+	NA	+	
68	CNEP_3_HU	+	NA	+	
69	CNEP_3_MX	+	NA	+	
70	CNEP_3_MZ	+	NA	+	
71	CNEP_3_PT	+	NA	+	
72	CNEP_3_TW	+	NA	+	

73	CNEP_3_UY	+	NA	+	
74	CNEP_3_ZA	+	NA	+	
75	CNEP_4_DO	+	NA	+	
76	CNEP_4_ES	+	NA	+	
77	CNEP_4_ID	+	NA	+	
78	CNEP_4_MX	+	NA	+	
79	CNEP_4_US	+	NA	+	
80	CNEP_4_ZA	+	NA	+	
81	EB_0_1	NA	+	+	
82	EB_0_2	NA	+	+	
83	EB_3	NA	+	+	
84	EB_5	NA	+	+	
85	EB_8	NA	+	+	
86	EB_9	NA	+	+	
87	EB_11	NA	+	+	
88	EB_13	NA	+	+	
89	EB_15	NA	+	+	
90	EB_18	NA	+	+	
91	EB_19	+	+	+	
92	EB_21	+	+	+	
93	EB_24	+	+	+	
94	EB_25	+	+	+	
95	EB_27	+	+	+	
96	EB_29	+	+	+	
97	EB_31	+	+	+	
98	EB_31A	+	+	+	
99	EB_34_1	+	+	+	

100	EB_35	+	+	+	
101	EB_40	+	+	+	
102	EB_42	+	+	+	
103	EB_44	+	+	+	
104	EB_44_2bis	+	+	+	
105	EB_47_1	+	+	+	
106	EB_49	+	+	+	
107	EB_52_1	+	+	+	
108	EB_54_1	+	+	+	
109	EB_55_1	+	+	+	
110	EB_56_1	+	+	+	
111	EB_57_1	NA	+	+	
112	EB_58_2	+	+	+	
113	EB_60_1	+	+	+	
114	EB_62_2	+	+	+	
115	EB_63_4	+	+	+	
116	EB_64_4	+	+	+	
117	EB_66_1	+	+	+	
118	EB_68_1	+	+	+	
119	EB_72_1	+	+	+	
120	EB_73_4	+	+	+	
121	EB_74_1	+	+	+	
122	EB_76_2	+	+	+	
123	EB_77_3	+	+	+	
124	EB_79_3	+	+	+	
125	EB_81_4	+	NA	+	
126	EB_83_3	+	NA	+	

127	EB_85_2	+	NA	+	
128	EB_87_3	+	NA	+	
129	EQLS_1_3	+	+	+	
130	ESS_1	+	NA	+	
131	ESS_2	+	NA	+	
132	ESS_3	+	NA	+	
133	ESS_4	+	NA	+	
134	ESS_5	+	NA	+	
135	ESS_6	+	NA	+	
136	ESS_7	+	NA	+	
137	ESS_8	+	NA	+	
138	EVS_1_4	NA	+	+	
139	ISJP_1_2	NA	+	+	
140	ISSP_1985	+	+	+	
141	ISSP_1986	+	+	+	
142	ISSP_1987	+	+	+	
143	ISSP_1988	+	+	+	
144	ISSP_1989	+	+	+	
145	ISSP_1990	+	+	+	
146	ISSP_1991	+	+	+	
147	ISSP_1992	+	+	+	
148	ISSP_1993	+	+	+	
149	ISSP_1994	+	+	+	
150	ISSP_1995	+	+	+	
151	ISSP_1996	+	+	+	
152	ISSP_1997	+	+	+	
153	ISSP_1998	+	+	+	

154	ISSP_1999	+	NA	+	
155	ISSP_2000	+	+	+	
156	ISSP_2001	+	+	+	
157	ISSP_2002	+	+	+	
158	ISSP_2003	+	+	+	
159	ISSP_2004	+	+	+	
160	ISSP_2005	+	+	+	
161	ISSP_2006	+	NA	+	
162	ISSP_2007	+	NA	+	
163	ISSP_2008	+	+	+	
164	ISSP_2009	+	+	+	
165	ISSP_2010	+	NA	+	
166	ISSP_2011	+	+	+	
167	ISSP_2012	+	+	+	
168	ISSP_2013	+	+	+	
169	ISSP_2014	+	+	+	
170	ISSP_2015	+	+	+	
171	LB_1995	+	NA	+	
172	LB_1996	+	NA	+	
173	LB_1997	+	NA	+	
174	LB_1998	+	NA	+	
175	LB_2000	+	NA	+	
176	LB_2001	+	NA	+	
177	LB_2002	+	NA	+	
178	LB_2003	+	NA	+	
179	LB_2004	+	NA	+	
180	LB_2005	+	NA	+	

181	LB_2006	+	NA	+	
182	LB_2007	+	NA	+	
183	LB_2008	+	NA	+	
184	LB_2009	+	NA	+	
185	LB_2010	+	NA	+	
186	LB_2011	+	NA	+	
187	LB_2013	+	NA	+	
188	LB_2015	+	NA	+	
189	LB_2016	+	NA	+	
190	LITS_1	+	NA	+	
191	LITS_2	+	NA	+	
192	LITS_3	+	NA	+	
193	NBB_1_6	NA	+	+	
194	NEB_1	+	+	+	questionnaire online
195	NEB_2	+	+	+	questionnaire online
196	NEB_3	+	+	+	questionnaire online
197	NEB_4	+	+	+	questionnaire online
198	NEB_5	+	+	+	questionnaire online
199	NEB_6	+	+	+	questionnaire online
200	NEB_7	+	+	+	questionnaire online
201	PA1	NA	+	+	
202	PA2	NA	+	+	
203	PPE7N_AT	NA	+	+	
204	PPE7N_IN	NA	+	+	
205	PPE7N_JP	NA	+	+	
206	PPE7N_NG	NA	+	+	
207	PPE7N_NL	NA	+	+	

208	PPE7N_US	NA	+	+	
209	PPE7N_YU	NA	+	+	
210	VPCPCE_CZ	+	NA	+	
211	VPCPCE_HU	+	NA	+	
212	VPCPCE_RU	+	NA	+	
213	VPCPCE_SK	+	NA	+	
214	VPCPCE_UA	+	NA	+	
215	WVS_1_6	+	+	+	wave level questionnaires