SDR2 Database Overview

The SDR2 Database consists of three datasets produced in the Survey Data Recycling project. These files are:

1. The MASTER file, which stores harmonized information for a total of 4,402,489 respondents. It includes:
   a. 50 harmonized target variables (T-prefixed variables)
   b. 74 harmonization control variables (C-prefixed variables), which capture source item properties and harmonization decisions. Harmonization controls generally do not vary across respondents of a given national survey. Exceptions are noted in specific target variables' General Variable Report, GVR.
   c. 7 source data quality controls (Q-prefixed variables), including flags for non-unique cases or cases related to suspicious values of source age and household size. These variables are measured at the respondent level (e.g., missing case ID).

2. The PLUG-SURVEY file is an auxiliary dataset containing controls for source data quality and a set of technical variables needed for merging this file with the MASTER file. Quality control variables on the national survey level pertain to two main aspects: (1) survey methodology and survey quality as reflected in documentation, and (2) general data quality (e.g., presence of non-unique records in a national survey, availability and correctness of survey weights). There are also quality control variables measured at the survey wave level, specifically focusing on processing error measures of discrepancies between data and documentation for selected variables.

3. The PLUG-COUNTRY file is a dictionary of countries and territories used in the MASTER file. It contains basic geographical information: alpha/numeric standardized ISO-3166 country codes, country names, and codes and names for micro- and macro-regions of the world.

Table 1 provides basic information about the SDR2 files, including their names as referenced throughout SDR2 documentation, key variables, and names of the corresponding documentation files. The key variables can be used either for relating individual respondents in the MASTER file and source data files (T_DATAFILE and T_ID), or for merging the MASTER file with the PLUG files. The link between MASTER's T_DATAFILE and the source data file can be established through the list of source data files in SDR2_Source_Data_Files.xlsx. T_ID is an SDR variable. It does not appear in source data files, but it refers to the position of a case (respondent) in the corresponding source file, i.e., it uniquely identifies a case (respondent) within a source data file.

Note that a national survey is identified uniquely by the combination of four variables, T_PROJECT_NAME, T_PROJECT_WAVE, T_COUNTRY_L1U, T_COUNTRY_L2U, which is conceptually

equivalent to project–year–country. The reason we use two country identifiers is the following. In a typical situation, a national survey is carried out in the whole country, e.g. Germany (DE). In such cases T_COUNTRY_L1U equals T_COUNTRY_L2U and either of them can be matched with T_COUNTRY from the PLUG_COUNTRY file. Sometimes, however, data producers distinguish parts of a country, in our example Eastern and Western Germany (DE-E and DE-W, respectively). This may be done for different reasons, including providing data only for a part of a country, providing weighting factors separately for parts of the countries, or carrying surveys separately in different parts of the countries. In such cases T_COUNTRY_L1U reflects the smallest country part in the source data (e.g. DE-E) while T_COUNTRY_L2U is a harmonized country code (e.g. DE). For additional information, see SDR2 MASTER File Overview.

The description of the three SDR2 datasets is provided in the following files: SDR2 MASTER File Overview, SDR2 PLUG_SURVEY File Overview and SDR2 PLUG_COUNTRY File Overview.

Table 1. Structure of the SDR2 Database: description of files and their contents

Name of file	**Content description**	**Key variables / case identifiers**	**Documentation files**
MASTER	Individual (respondent) level data including target, harmonization control and quality control variables	T_DATAFILE T_ID T_PROJECT_NAME T_PROJECT_WAVE T_COUNTRY_L1U T_COUNTRY_L2U	SDR2_MASTER_File_Overview.pdf SDR2_Missing_Codes_Schema.pdf SDR2_Source_Data_Files.xlsx SDR2_Cotton_File.xlsm SDR2_Metadata.xlsx SDR2_MASTER_Syntax_File.sql General Target Variable reports (GVR) pdf files Detailed Target Variable reports (DVR) xlsx files CrossWalk Table reports (CWT) xlsx files for target variable of SDR2 (see Appendix for the detailed file list)
PLUG-COUNTRY	Country-level data including ISO-3166 country codes and names of geographical regions	T_COUNTRY	SDR2_PLUG_COUNTRY_File_Overview.pdf SDR2_Source_Data_Files.xlsx
PLUG-SURVEY	Characteristics of national surveys including sampling method and response rate, indicators of survey quality, availability and correctness of survey weights, and discrepancies between data and documentation	T_PROJECT_NAME T_PROJECT_WAVE T_COUNTRY_L1U T_COUNTRY_L2U	SDR2_PLUG_SURVEY_File_Overview.pdf SDR2_Source_Data_Files.xlsx

Researchers can access the three SDR2 datasets (in SPSS and STATA formats) and the full SDR2 documentation on Harvard Dataverse. The SDR2 data are also available in CSV format (with corresponding syntax for reading the files into STATA/SPSS) from the SDR Portal (http://cse-dnc215869s.cse.ohio-state.edu/), whose extraction utility offers the possibility to download them in full or customize data selection according to specific variables, survey projects, countries and years.

Appendix. Master file: List of documentation reports accompanying SDR2 target variables.

1. T_AGE_BIRTHYR_DVR_SDR2.xlsx
2. T_AGE_BIRTHYR_GVR_SDR2.pdf
3. T_AGE_CWT_SDR2.xlsx
4. T_BIRTHYR_CWT_SDR2.xlsx
5. T_BOYCOTT_CWT_SDR2.xlsx
6. T_BOYCOTT_DVR_SDR2.xlsx
7. T_BOYCOTT_GVR_SDR2.pdf
8. T_CIVIL_STATUS_CWT_SDR2.xlsx
9. T_CIVIL_STATUS_DVR_SDR2.xlsx
10. T_CIVIL_STATUS_GVR_SDR2.pdf
11. T_DEMONST_CWT_SDR2.xlsx
12. T_DEMONST_DVR_SDR2.xlsx
13. T_DEMONST_GVR_SDR2.pdf
14. T_DEMSAT_CWT_SDR2.xlsx
15. T_DEMSAT_DVR_SDR2.xlsx
16. T_DEMSAT_GVR_SDR2.docx
17. T_EDU_CATEGORY_OPEN_CWT_SDR2.xlsx
18. T_EDU_CWT_SDR2.xlsx
19. T_EDU_DVR_SDR2.xlsx
20. T_EDU_GVR_SDR2.pdf
21. T_EDU_VOCATIONAL_CWT_SDR2.xlsx
22. T_EDU_YEARS_CWT_SDR2.xlsx
23. T_EDU_YEARS_DVR_SDR2.xlsx
24. T_EDU_YEARS_GVR_SDR2.pdf
25. T_GENDER_CWT_SDR2.xlsx
26. T_GENDER_DVR_SDR2.xlsx
27. T_GENDER_GVR_SDR2.pdf
28. T_HAPPY_CWT_SDR2.xlsx
29. T_HAPPY_GVR_SDR2.pdf
30. T_HAPPY_LIFESAT_DVRS_SDR2.xlsx
31. T_HEALTH_CWT_SDR2.xlsx
32. T_HEALTH_DVR_SDR2.xlsx
33. T_HEALTH_GVR_SDR2.pdf
34. T_HH_SIZE_CWT_SDR2.xlsx
35. T_HH_SIZE_GVR_SDR2.pdf
36. T_INCOME_CWT_SDR2.xlsx

37. T_INCOME_DVR_SDR2.xlsx
38. T_INCOME_GVR_SDR2.pdf
39. T_INT_YEAR_CWT_SDR2.xlsx
40. T_INTPOL_CWT_SDR2.xlsx
41. T_INTPOL_DVR_SDR2.xlsx
42. T_INTPOL_GVR_SDR2.pdf
43. T_LIFESAT_CWT_SDR2.xlsx
44. T_LIFESAT_GVR_SDR2.pdf
45. T_MEMBER_ENVIRON_GVR_SDR2.pdf
46. T_MEMBER_ENVIRON_CWT_SDR2.xlsx
47. T_MEMBER_ENVIRON_DVR_SDR2.xlsx
48. T_MEMBER_ORG_CWT_SDR2.xlsx
49. T_MEMBER_ORG_DVR_SDR2.xlsx
50. T_MEMBER_ORG_GVR_SDR2.pdf
51. T_MEMBER_PARTY_CWT_SDR2.xlsx
52. T_MEMBER_PARTY_DVR_SDR2.xlsx
53. T_MEMBER_PARTY_GVR_SDR2.pdf
54. T_MEMBER_RELIG_CWT_SDR2.xlsx
55. T_MEMBER_RELIG_DVR_SDR2.xlsx
56. T_MEMBER_RELIG_GVR_SDR2.pdf
57. T_MEMBER_UNION_CWT_SDR2.xlsx
58. T_MEMBER_UNION_DVR_SDR2.xlsx
59. T_MEMBER_UNION_GVR_SDR2.pdf
60. T_NEVER_DISCPOL_CWT_SDR2.xlsx
61. T_NEVER_DISCPOL_DVR_SDR2.xlsx
62. T_NEVER_DISCPOL_GVR_SDR2.pdf
63. T_PETITION_CWT_SDR2.xlsx
64. T_PETITION_DVR_SDR2.xlsx
65. T_PETITION_GVR_SDR2.pdf
66. T_RURAL_CWT_SDR2.xlsx
67. T_RURAL_DVR_SDR2.xlsx
68. T_RURAL_EXTEND_CWT_SDR2.xlsx
69. T_RURAL_GVR_SDR2.docx
70. T_SETTLE_SIZE_CWT_SDR2.xlsx
71. T_SETTLE_SIZE_GVR_SDR2.pdf
72. T_TRGOV_CWT_SDR2.xlsx
73. T_TRGOV_GVR_SDR2.pdf

74. T_TRLEG_CWT_SDR2.xlsx
75. T_TRLEG_GVR_SDR2.pdf
76. T_TRPARL_CWT_SDR2.xlsx
77. T_TRPARL_GVR_SDR2.pdf
78. T_TRPARTY_CWT_SDR2.xlsx
79. T_TRPARTY_GVR_SDR2.pdf
80. T_TRPEOPLE_CWT_SDR2.xlsx
81. T_TRPEOPLE_GVR_SDR2.pdf
82. T_TRUSTS_DVRS_SDR2.xlsx
83. T_VOTED_CWT_SDR2.xlsx
84. T_VOTED_DVR_SDR2.xlsx
85. T_VOTED_GVR_SDR2.pdf
86. T_WEIGHT_L2U_GVR_SDR2.pdf
87. T_WEIGHT_L2U_CWT_SDR2.xlsx