

# Quality Check & General Variable Report

Weronika Boruc

[wboruc@ifispan.waw.pl](mailto:wboruc@ifispan.waw.pl)

Institute of Philosophy and Sociology, Polish Academy of Sciences

# Survey Harmonization ex-post: WORKFLOW

Creating harmonized measures *ex-post* **strongly relies** on released survey data and documentation linked to it.

1. Theoretical model and concept specification
2. Criteria for data selection -> Time zero: data and documentation downloaded
3. Source variables: first (inclusive) selection; item availability template
4. (Re)Define target variable based on source variables; final source variable list
5. Creating harmonization control variables
6. Cross-walk coding and additional source variables check
7. Quality control variables for items, documentation and data
8. Final revisions of harmonization and its documentation; general variable report

# 7. Additional quality check: Processing Errors Report (PER)

Consistency of **Documentation** Information (Questionnaires, Codebooks) with **Data** Information (Data Dictionaries, Realized Data);

Why?

- Such inconsistencies influence the quality of data.

- ✓ Age
- ✓ Year of Birth
- ✓ Gender
- ✓ Education Levels
- ✓ Education in Years
- ✓ Participation in Demonstrations
- ✓ Trust in Parliament

## B. TRUST IN INSTITUTIONS

7-18. I'M GOING TO NAME A NUMBER OF INSTITUTIONS. FOR EACH ONE, PLEASE TELL ME HOW MUCH TRUST YOU HAVE IN THEM. IS IT A GREAT DEAL OF TRUST, QUITE A LOT OF TRUST, NOT VERY MUCH TRUST, OR NONE AT ALL?

Qs. 7-18.

(RING CARDS – SHOWCARD)

	A Great Deal of Trust	Quite a Lot of Trust	Not Very Much Trust	None At All
--	-----------------------	----------------------	---------------------	-------------

7.	The Courts	1	2	3	4
8.	The national government [in capital city]	<b>1 = A Great Deal of Trust</b>			
9.	Political parties [not any specific party]				
10.	Parliament	1	2	3	4
11.	Civil service	1	2	3	4
12.	The military	1	2	3	4
13.	The police <Optional>	1	2	3	4
14.	Local government	1	2	3	4
15.	Newspapers <Optional>	1	2	3	4

q007 '7. How much trust do you have in the courts?'

q008 '8. How much trust do you have in the national government?'

q009 '9. How much trust do you have in political parties?'

q010 '10. How much trust do you have in Parliament?'

q011 '11. How much trust do you have in civil service?'

q012 '12. How much trust do you have in the military?'

1 'None at all'

2 'Not very much trust'

3 'Quite a lot of trust'

4 'A great deal of trust'

98 'Don't know'

99 'No Answer'

**1 = None at all**

	A	B	C	E
1	dataset	variable's name	variable's label (Google translation from Spanish)	values
40	ABS_1	q007	How much trust do you have in the courts?	0 Not sure 1 None at all 2 Not very much trust 3 Quite a lot of trust 4 A great deal of trust 98 Don't know 99 No Answer
41	ABS_1	q008	How much trust do you have in the national government?	0 Not sure 1 None at all 2 Not very much trust 3 Quite a lot of trust 4 A great deal of trust 98 Don't know 99 No Answer
42	ABS_1	q009	How much trust do you have in political parties?	0 Not sure 1 None at all 2 Not very much trust 3 Quite a lot of trust 4 A great deal of trust 98 Don't know 99 No Answer
43	ABS_1	q010	How much trust do you have in Parliament?	0 Not sure 1 None at all 2 Not very much trust 3 Quite a lot of trust 4 A great deal of trust 98 Don't know 99 No Answer

**1 = None at all**

# Sources of Information & Tools:

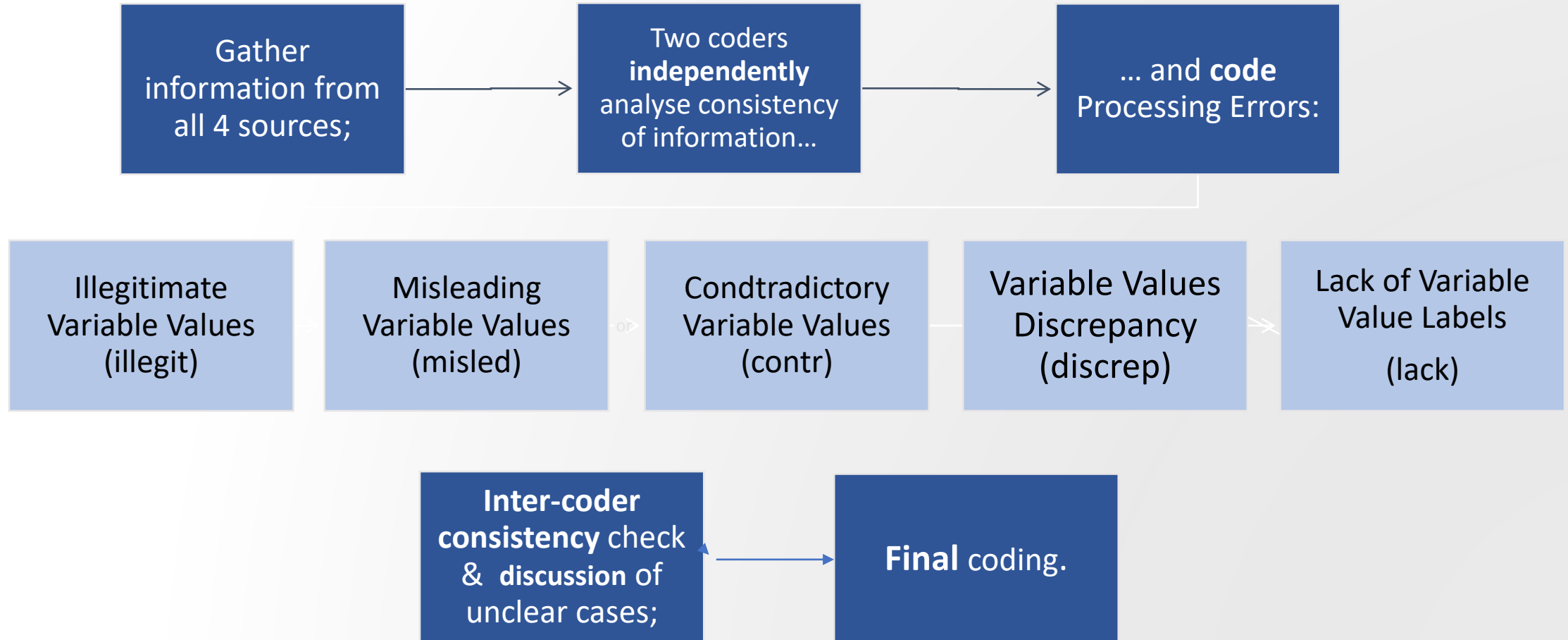
For quality check we need **information** from **four** different **sources**:

- Questionnaire
- Codebook
- Data dictionary
- Realized data.

We use **Detailed Variable Report** files + add columns for coding **Processing Errors**;

As result:  
**1 file** with all information about source variable: **DVR + PER**.

# Methodology of Coding Processing Errors



What do we  
do with PER  
information?

Update General Variable Reports  
(especially the „Special Cases” section);

Provide SDR database users with  
information about Documentation/Data  
inconsistency: Index on survey-wave level;

Flags for Target Variables about the quality  
of source variables (in DVR-PER – and in  
SDR 2.0 Data File).

# Problems encountered & lessons learned

While coding, we have different ideas about both the presence & the type of Processing Errors, and it's difficult to remember our reasons and arguments.

-> That's why we make comments about coding rationale.

The Processing Errors we encounter are of different level of seriousness.

-> We attach different 'weight' to different types of Errors (e.g. „lack” is not as serious as „discrep”)



How PER can  
be useful for  
you?

**SDR data users:** Information concerning quality of source data.

**Other harmonization teams:** Templates for gathering information needed for checking documentation/data inconsistencies

# Survey Harmonization ex-post: WORKFLOW

Creating harmonized measures *ex-post* **strongly relies** on released survey data and documentation linked to it.

1. Theoretical model and concept specification
2. Criteria for data selection -> Time zero: data and documentation downloaded
3. Source variables: first (inclusive) selection; item availability template
4. (Re)Define target variable based on source variables; final source variable list
5. Creating harmonization control variables
6. Cross-walk coding and additional source variables check
7. Quality control variables for items, documentation and data
8. Final revisions of harmonization and its documentation; general variable report

## 8. Final revisions of harmonization and its documentation and General Variable Report

For **each Target Variable** in SDR2.0 we prepare:

- **Detailed Variable Report (DVR)**  
(in some cases + **Processing Errors Report**)
- **Cross-Walk Table (CWT)**
- **General Variable Report (GVR)**
- **Harmonization Syntax**

## 8. cd. General Variable Report (GVR)

Structured information about:

- **Target Variable** (clarification of Target Concept and its operationalization);
- **Source Variables** (availability within project/waves/years/countries, specificities, most common question wordings);
- **Harmonization rules & procedures;**
- **Control Variables** (controlling specific features of source variables and/or survey quality).
- **Special cases** (highlighting difficulties and transparent reasoning for our decisions).