

CSDI 2015 Workshop, London, March 26-28

**Re-Thinking *Ex Post* Harmonization for
Comparative Research:
Data Recycling of 22 International Survey Projects**

**Kazimierz M. Slomczynski, CONSIRT
Irina Tomescu-Durbow, CONSIRT
J. Craig Jenkins, The Ohio State University (OSU)**

**Cross-national Studies, Interdisciplinary Research and Training Program - CONSIRT,
the Polish Academy of Sciences and OSU
consirt.osu.edu**

**Research supported by the Polish National Science Centre (2012/06/M/HS6/00322)
as part of the study *Democratic Values and Protest Behavior: Data Harmonization,
Measurement Comparability, and Multi-Level Modeling*
(hereafter, Harmonization Project)**

Criteria of selecting survey projects

Projects:

- containing questions about political attitudes and behaviors**
- designed as cross-national, and, preferably, multi-wave;**
- with the samples intended as representative of the adult population of given country or territory;**
- non-commercial;**
- freely available in the public domain;**
- with documentation - study description, codebook and/or questionnaire - in English**

Abbrev.	Survey Project	Time span	Waves	Files	Data Sets	Cases
			Counts			
AFB	Afrobarometer	1999-2009	4	4	66	98942
AMB	Americas Barometer	2004-2012	5	1	92	151341
ARB	Arab Barometer	2006-2011	2	2	16	19684
ASB	Asian Barometer	2001-2011	3	3	30	43691
ASES	Asia Europe Survey	2000	1	1	18	18253
CB	Caucasus Barometer	2009-2012	4	4	12	24621
CDCEE	Consolidation of Democracy (C./East Europe)	1990-2001	2	1	27	28926
CNEP	Comparative National Elections Project	2004-2006	1	8	8	13372
EB	Eurobarometer	1983-2012	7	7	152	138753
EQLS	European Quality of Life Survey	2003-2012	3	1	93	105527
ESS	European Social Survey	2002-2013	6	2	146	281496
EVS/WVS	European Values Study/World Values Survey	1981-2009	9	1	312	423084
ISJP	International Social Justice Project	1991-1996	2	1	21	25805
ISSP	International Social Survey Programme	1985-2013	13	13	363	493243
LB	Latinobarometro	1995-2010	15	15	260	294965
LITS	Life in Transition Survey	2006-2010	2	2	64	67866
NBB	New Baltic Barometer	1993-2004	6	1	18	21601
PA2	Political Action II	1979-1981	1	1	3	4057
PA8NS	Political Action – 8 Nation Study	1973-1976	1	1	8	12588
PPE7N	Political Participation and Equality in 7 Nations	1966-1971	1	7	7	16522
VPCPCE	Values/Political Change, Postcommunist E	1993	1	5	5	4723
Total		1966-2013	89	81	1721	2289060

Data

- **1,721 national surveys with over 95% of samples from 500 to 3,000 respondents**
- **National surveys conducted in 142 countries and territories over a period of almost 50 years**
- **All these surveys contain over 25,000 variables**
- **From 150 to 200 variables are identical or similar in large majority of 89 waves**
- **From 25 to 40 variables deal directly with political attitudes and behaviors**

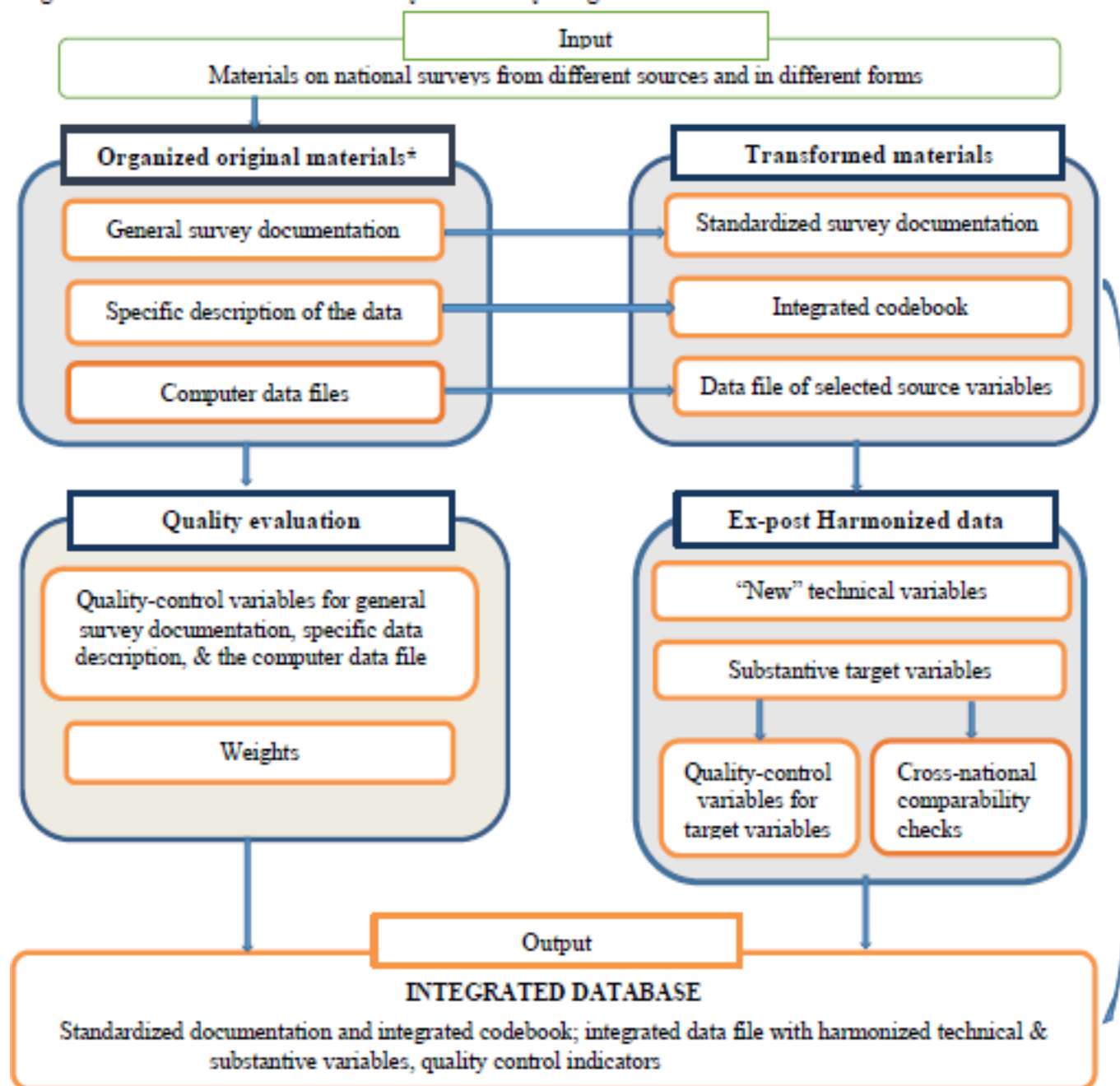
Survey Data Recycling (SDR)

- SDR = framework for (re)-processing cross-national survey data;
- SDR concerns survey data quality control & *ex-post* survey harmonization, to:
 - (a) account for “messiness” of the original source data,
 - (b) provide “comparable” data

Thus, SDR expands the scope of extant projects (time, space, number of observations, types of indicators)

Unifying thesis: account for errors & biases in original surveys & harmonization procedures *via* different types of quality control variables, to be included in substantive analyses.

Figure 1. General Schema of Survey Data Recycling



Formal framework

1. Relationship between target variables T and source variables S

$$T = f(S) \text{ [substantive decision of f]}$$

2. Relationship between T and X and two types of control variables: Q and H

X = Substantive independent variables

Q = Data Quality Controls for (a) survey documentation, (b) specific data description, and (c) data in the computer files

H = Harmonization Quality Controls of specific procedures that could influence validity and reliability of T

$$T = b_0 + b_1 Q + b_2 H + b_3 X + e$$

Explanation

T = Target Variable

Q = Data Quality Controls

H = Harmonization Quality Controls

X = Substantive Independent Variables

$$T = b_0 + b_1 Q + b_2 H + b_3 X + e$$

If e is negligible & $b_1, b_2, b_3 = 0$, $b_0 = f(S)$ for $T = f(S)$

If b_1 and/or b_2 unequal 0, some intervention may be needed to correct for errors and biases in T .

Minimal solution: partialing out effects of Q and H in the estimated impact of X on T .

Survey quality-control Indicators

- (a) survey documentation;**
- (b) (in)consistency between data description (e.g. in codebooks, questionnaires) & data records and data records in computer file;**
- (c) computer data records themselves.**

General Survey Documentation: How the data has been collected?	Answers
Does the survey documentation specify the type of <u>sample used</u>?	Yes = 0 No = 1
Does the survey documentation provide information on the <u>response rate</u>?	Yes = 0 No = 1
Was the <u>questionnaire</u> back-translated or translation checked in some other way?	Yes = 0 No = 1
Is there any evidence that the <u>questionnaire</u> was pre-tested?	Yes = 0 No = 1
Does the documentation show that the <u>fieldwork</u> was controlled?	Yes = 0 No = 1
Effect of negative answers (No = 1) : Reduction of confidence in the data	

<u>Specific Data Description: How have the data been defined?</u>	Answers
Do variable values in the codebook correspond to values in the data file?	Yes = 0 No = 1
Eight binary variables describing discrepancies between data description and the data file	
(Ilona and Olena)	
Effect of negative answers (No = 1): Decrease of interpretability of the data	

Computer Data File: Are the data formally correct?	Answers
Do survey cases (respondents) have unique identification numbers (IDs)?	Yes = 0 No = 1
Are survey weights free of formal errors? (Marcin and Przemek)	Yes = 0 No = 1
Is the proportion of missing values for gender and age within the standard limits (< 5%)?	Yes = 0 No = 1
Is the data file free from repeated cases (duplicates)?	Yes = 0 No = 1
Effect of negative answers (No = 1) : Possible distortion of the research results based on the data	

Data Harmonization Controls

Variables pertaining to:

- **Wording**
- **Scales**
- **Item Context**

[Measurement properties]

(Marta paper on wording, scales, and item context)

Harmonization Project Team

- **Kazimierz M. Słomczyński**, Polish Academy of Sciences
and Ohio State University
- **Irina Tomescu-Dubrow**, CONSIRT
- **Craig Jenkins**, Ohio State University
- **Marta Kołczyńska**, Ohio State University
- **Ilona Wysmułek**, Graduate School for Social Research (Warsaw)
- **Olena Oleksiyenko**, Graduate School for Social Research (Warsaw)
- **Przemek Powańko**, Polish Academy of Sciences
- **Marcin Zieliński**, University of Warsaw
- **Matt Schoene**, Ohio State University

and

Joshua Dubrow, **Anna Franczak**, **Anastas Vangeli**, **Jakub Wysmulek** - Graduate School for Social Research (Warsaw)