



Research & Methods
ISSN 1234-9224 Vol. 24 (1, 2015): 49–89
The Ohio State University
Columbus, Ohio, USA
Institute of Philosophy and Sociology
Polish Academy of Sciences, Warsaw, Poland
www.askresearchandmethods.org

Harmonization: Newsletter on Survey Data Harmonization in the Social Sciences

EDITORS' NOTE

The following are reprints of the articles that appeared this year in the Newsletter, *Harmonization: Newsletter on Survey Data Harmonization in the Social Sciences*. This Newsletter is a production of, and was originally published by, Cross-national Studies: Interdisciplinary Research and Training program (CONSIRT.osu.edu), The Ohio State University and the Polish Academy of Sciences (PAN). It was edited by Irina Tomescu-Dubrow and Joshua Kjerulf Dubrow, CONSIRT. Marta Kołczyńska of The Ohio State University provided technical assistance. The first issue appeared as Volume 1, Number 1, in Winter 2015. The second appeared as Volume 1, Number 2, in Fall 2015. They were first published in the website, consirt.osu.edu/newsletter.

The catalyst for the Newsletter is the project, “Democratic Values and Protest Behavior: Data Harmonization, Measurement Comparability, and Multi-Level Modeling” (hereafter, Harmonization Project). Financed by the Polish National Science Centre in the framework of the Harmonia grant competition (2012/06/M/HS6/00322), the Harmonization Project joins the Institute of Philosophy and Sociology PAN and the OSU Mershon Center for International Security Studies in creating comparable measurements of political protest, social values, and demographics using information from well-known international survey projects. The team includes: Kazimierz M. Slomczynski (PI), J. Craig Jenkins (PI), Irina Tomescu-Dubrow, Joshua Kjerulf Dubrow, Przemek Powalko, Marcin W. Zieliński, and research assistants: Marta Kołczyńska, Matthew Schoene, Ilona Wyszumłek, Olena Oleksiyenko, Anastas Vangeli, and Anna Franczak. For more information about the harmonization project, please visit dataharmonization.org.

All volumes of *Ask: Research and Methods*, 1995 to the present, including each individual article, are permanently archived in The Ohio State University's Knowledge Bank. You can find more about it at the website, CONISRT.osu.edu/askresearchandmethods.

WELCOME!

by Joshua Kjerulf Dubrow and Irina Tomescu-Dubrow
CONSIRT

Welcome to the first issue of *Harmonization: Newsletter on Survey Data Harmonization in the Social Sciences*. Survey data harmonization and big data are innovative forces that are leading to new, emergent and interdisciplinary knowledge across the social sciences. The purposes of this newsletter are to share news and communicate with the growing community of scholars, institutions and government agencies who work on harmonizing social survey data and other projects with similar focus.

We pay special attention to the methodology of survey data harmonization. We intend for this newsletter to contribute to the development of international research and standards on methodological issues such as data comparability, data quality, proper data documentation, and data storage and access, as well as analytical procedures that can contend with the demands of harmonized data.

A Brief History of Survey Data Harmonization Projects

by Joshua Kjerulf Dubrow

Polish Academy of Sciences and CONSIRT

This article gives a brief overview of ex post cross-national survey data harmonization (SDH) projects in the social sciences from the 1980s to the 2010s (see also Burkhauser and Lillard 2005; Granda, Wolf, and Hadorn 2010; Dubrow and Tomescu-Dubrow 2014).

There are two major types of SDH projects. One are large scale projects designed to produce data on a range of research topics with open research questions. They involve multiple institutions – including governments, and especially their financing – and large numbers of researchers and assistants. These projects produce harmonized data and corresponding user manuals, as well as publications on the use of these data for addressing substantive issues. The second type are projects designed by small research teams to answer specific pre-determined research questions. Here, harmonization is limited to the variables needed to answer the research questions. This article focuses on large-scale projects.

One of the earliest attempts to integrate data from different extant surveys, and perhaps the most successful, is the Luxembourg Income Study, now simply called LIS. The idea of LIS was generated by a conference on the topic of poverty in cross-national perspective, held in Luxembourg in 1982 (for a detailed history, see Smeeding, Schmaus, and Allegrezza 1985: 2–4).

While LIS was getting off the ground, scholars interested in the concept of “time use” also started to consider how to compare all of the Time Use Studies (TUS) conducted in various countries, past and present. The resulting project, named the Multinational Time Use Study (MTUS), has its roots in the 1970s, but only took shape as a harmonized time use study in the 1980s (for a detailed history, see MTUS User’s Guide 2013: Chapter 2). MTUS is based on time use diaries. The European Foundation for the Improvement of Living and Working Conditions (EFILWC), an agency of the European Union, paid for the initial release of MTUS; the collaboration between MTUS researchers and the EU led to the Harmonized European Time Use Study, or HETUS.

One of the most significant SDH projects initiated in the 1990s is the Cross-national Equivalent File (CNEF) (see Lillard’s article in this Newsletter). CNEF is simultaneously based on the successful LIS model and designed to be different from LIS. Unlike LIS, CNEF harmonizes household panel studies and was designed to be developed and enhanced by its user community. CNEF can

be called a bottom-up approach, with users having strong say in the direction of CNEF's target variables, in contrast to LIS' top-down approach. When it comes to top-down or bottom-up in SDH, there are no ideal solutions, as LIS uses its working papers to understand how users employ the data.

The early 2000s saw the maturation of LIS, CNEF and HETUS, and the creation of new SDH projects. An early project was the Consortium of Household Panels for European Socio-economic Research (CHER). CHER was initially funded by the European Commission for over one million Euros between 2000 and 2003, and coordinated by Centre de Recherche en Sciences Sociales (CEPS), a research bureau in Luxembourg. CHER is substantively similar in its harmonization aims to CNEF, namely the harmonization of already collected panel data. By 2003, CHER had data going back to the 1980s. CHER ended in 2003, and was not updated since.

The European Union Statistics on Income and Living Conditions (EU-SILC) was formally created in 2004 and is run by Eurostat. Like CNEF and CHER, EU-SILC deals with ex-post harmonized data of coordinated larger-scale surveys; it includes cross-sectional and longitudinal surveys on income, poverty, social exclusion and living conditions in the European Union.

The two largest ex-post cross-national SDH projects run by political scientists and sociologists are the Global Barometer Survey [GBS] and the International Stratification and Mobility File [ISMF]. A team of political scientists constructed GBS, which spans 2003-2006, from the surveys of 55 different countries or regions on democratic politics. The GBS, according to Bratton (2009: 1), is "based on stand-alone barometer surveys for various world regions in East Asia, South Asia, Latin America, sub-Saharan Africa, and the Arab Middle East."

Survey data harmonization has no institutionalized apparatus: no journal, no professional association, no academic department, and no research center; it does not even have a separate handbook.

ISMF allows researchers to compare social stratification and social mobility patterns across countries and time. By 2009, ISMF expanded to over 250 surveys from 52 nations, with some surveys dating back to the 1940s. Its focus is on educational and occupational status of both respondents and their parents, and has harmonized demographics, education, employment status, occupation and income.

The 2010s have seen the continuation of CNEF, EU-SILC, and ISMF, as well as MTUS and HETUS. In 2013, the Harmonization Project joined the group of large-scale SDH projects. It is led by sociologists Kazimierz M. Slomczynski of the Polish Academy of Sciences and The Ohio State University, and J. Craig Jenkins, who represents the OSU Mershon Center for International Security Studies. The Harmonization Project focuses on political protest and its micro and macro-determinants, while also keeping the possibility of harmonizing variables relevant to other topics open. This newsletter features a description of the study.

Lessons from History

Regarding the history of social science SDH projects since the 1980s, there is evidence that these projects learn from each other: a new methodological field emerges. Yet, this field emerges without a coordinated effort to build a comprehensive theoretical and methodological base. One reason is that SDH has no institutionalized apparatus: no journal, no professional association, no academic department, and no research center; SDH does not even have a separate handbook. It is only in the last fifteen years that, in the social sciences, there is some attempt at a theory of SDH and the development of an appropriate methodology. Exemplary work in this regard are that of Hoffmeyer-Zlotnik and Wolf (2003), Minkel (2004), Granda and Blasczyk (2010), Granda, Wolf and Hadorn (2010). The Harmonization Project has recognized these achievements, and is addressing the problems already raised by pushing for a theory of data harmonization and by focusing on methodological issues.

REFERENCES

- Bratton, Michael. 2009. "Democratic Attitudes and Political Participation: An Exploratory Comparison across World Regions." Paper presented at the Congress of the International Political Science Association, Santiago, Chile, July.
- Burkhauser, Richard V. and Dean R. Lillard. 2005. "The Contribution and Potential of Data Harmonization for Cross-National Comparative Research." DIW Diskussionspapiere, No. 486.
- Granda, Peter and Emily Blasczyk. 2010. Data Harmonization. Guidelines for Best Practices in Cross-Cultural Surveys. Ann Arbor, MI: Survey Research Center at the Institute for Social Research, University of Michigan.
- Granda, Peter, Christof Wolf and Reto Hadorn. 2010. "Harmonizing Survey Data." Pp. 315-334 in Janet A. Harkness, Michael Braun, Brad Edwards, Timothy P. Johnson, Lars Lyberg, Peter Ph. Mohler, Beth-Ellen Pennell and Tom W. Smith (eds.) Survey Methods in Multinational, Multiregional, and Multicultural Contexts. New York: Wiley.
- Hoffmeyer-Zlotnik, Jurgen H. P. and Christof Wolf. 2003. "Comparing Demographic and Socio-Economic Variables across Nations: Synthesis and Recommendations." Pp. 389-406 in Jurgen H. P. Hoffmeyer-Zlotnik and Christof Wolf (ds.) Advances in Cross-national Comparison: A European Working Book for Demographic and Socio-Economic Variables. New York: Springer Science+Business Media.
- Minkel, Hartmut. 2004. Report on Data Conversion Methodology. CHINTEX Working Paper no. 20.
- Multinational Time Use Study. 2013. User's Guide and Documentation. timeuse.org
- Smeeding, Timothy, Günther Schmaus, and Serge Allegrezza. 1985. "An Introduction to LIS." Luxembourg Income Study Working Paper Series. Working Paper No. 1 (June).

Joshua Kjerulf Dubrow is Associate Professor at the Institute of Philosophy and Sociology, Polish Academy of Sciences and Projects and Labs Coordinator at CONSIRT. His edited book, *Political Inequality in an Age of Democracy: Cross-national Perspectives* was published by Routledge in 2014.

The Cross-National Equivalent File (CNEF): Harmonized Panel Survey Data in Eight Countries

Dean Lillard

The Ohio State University

The Cross-National Equivalent File (CNEF), a cooperative effort of individuals and institutions that collect panel survey data in (as of 2012) eight countries: the Panel Study of Income Dynamics (PSID) for the United States; the Socio-Economic Panel Study (SOEP) for Germany; the British Household Panel Survey (BHPS) for Great Britain; the Survey of Labour and Income Dynamics (SLID) for Canada; the Household, Income and Labour Dynamics in Australia (HILDA) Survey for Australia; the Korea Labor and Income Panel Study (KLIPS); the Russia Longitudinal Monitoring Study of the Higher School of Economics (RLMS-HSE); the Swiss Household Panel (SHP) for Switzerland.¹ In 2015, researchers at Keio University in Japan will contribute data files from two ongoing household panel studies – the Keio Household Panel Survey (KHPS) and the Japan Household Panel Survey (JHPS).

The CNEF project harmonizes data common to two or more of the country-based surveys, allows researchers access to both the harmonized and original data, provides all harmonization algorithms to interested researchers, and focuses on some of the most successful nationally representative ongoing longitudinal micro-data sets in the world.²

The CNEF differs from other standardization projects not only because it includes data from ongoing panel studies, but also because the development and expansion of the variable set is largely driven by research questions. The project adds equivalently defined variables when researchers develop cross-nationally comparable measures as part of a particular research project. Because those researchers are experts on the topic of their study, they not only inform themselves of specific country institutions, but they also bring their topic-specific expertise to bear. Consequently, the harmonized data included in the CNEF are an amalgam of the knowledge of many researchers answering a diverse set of questions. Just as importantly, the CNEF continuously evolves as researchers refine and add to the set of harmonized variables.

The CNEF is also distinguished by its inclusion of data on the same person over many years. These longitudinal data make it possible for cross-national researchers to use more powerful statistical methods to better control for otherwise unobserved person-specific heterogeneity in behavior. Furthermore, these panels allow researchers to exploit policy variation not only across countries but also

over time; variation that yields a richer understanding of human behavior. Finally, the design of each country's survey allows researchers to follow families across multiple generations. Consequently, the CNEF is increasingly used to study, from a cross-national as well as a cross-disciplinary perspective, how socio-economic status is correlated and transmitted across multiple generations (e.g. Butz and Torrey 2006).

NOTES

- 1 The CNEF is administered at Ohio State University in close collaboration with researchers at the Socio-Economic Panel Study at the German Institute for Economic Research (DIW Berlin) in Berlin, the Institute for Social and Economic Research (ISER) at the University of Essex, Statistics Canada in Ottawa, the Survey Research Center at the University of Michigan, the Melbourne Institute of Applied Economic and Social Research at the University of Melbourne, the University of Neuchâtel, the Center for Labor Policy Analysis at the Korea Labor Institute, at Demoscope (Moscow) and the Higher School of Economics in Russia, and at Keio University, Tokyo. For description of the project, see Burkhauser and Lillard 2005 and 2006; Frick, Jenkins, Lillard, Lipps, and Wooden 2007.
- 2 For more information, contact CNEF@osu.edu or visit the home page at <http://cnef.ehe.osu.edu/>

REFERENCES

- Butz, William P. and Barbara Boyle Torrey. 2006. "Some Frontiers in Social Science." *Science* 312: 1898–1900.
- Burkhauser, Richard V. and Dean R. Lillard. 2005. "The Contribution and Potential of Data Harmonization for Cross-National Comparative Research." *Journal of Comparative Policy Analysis* 7(4): 313-330.
- Burkhauser, Richard V. and Dean Lillard. 2006. "Evaluation of the Cross-National Comparability of the Survey of Health, Ageing, and Retirement in Europe, the Health and Retirement Study, and the English Longitudinal Study of Ageing." Report prepared for the National Institute on Ageing, Washington D.C.
- Frick, Joachim R., Stephen P. Jenkins, Dean R. Lillard, Oliver Lipps, and Mark Wooden. 2007. "The Cross-National Equivalent File (CNEF) and its Member Country Household Panel Studies." *Schmollers Jahrbuch - Journal of Applied Social Science Studies* 127(4): 627-654.

Dean Lillard is an Associate Professor in the Department of Human Sciences at The Ohio State University, and he directs and manages the Cross-national Equivalent File (CNEF). In 2015, Oxford University Press will publish his co-edited book, *Life-Course Smoking Behavior: Patterns and National Context in Ten Countries*, the product of a five-year NIH-funded project. He is an advisor for the Harmonization Project.

The Harmonization Project: Democratic Values and Protest Behavior in 22 International Survey Projects

by Irina Tomescu-Dubrow and Kazimierz M. Slomczynski

Polish Academy of Sciences and CONSIRT

The Democratic Values and Protest Behavior: Data Harmonization, Measurement Comparability, and Multi-Level Modeling study is financed by the (Polish) National Centre of Science and supported by The Ohio State University. CONSIRT hosts the project in Poland. While there are a number of survey data harmonization projects that have informed our own, each with their own acronyms (Dubrow and Tomescu-Dubrow 2014), we have come to call this large-scale research, simply, the Harmonization Project.

The Harmonization Project engages with the relationship between democracy and protest behavior in comparative, cross-national perspective.

Substantively, the project engages with the relationship between democracy and protest behavior in comparative, cross-national perspective. Political protest can be of various types, such as participation in demonstrations, signing petitions, or contacting politicians. Drawing on extant research (Benson and Rochon 2004; Kriesi 2004; Dubrow, Slomczynski and Tomescu-Dubrow 2008; Dalton Sickle and Weldon 2009; Marien, Hooghe and Quintelier 2010; Vrablikova 2013), we develop a two-level model where protest (individual-level) is explained by a set of theoretically-informed characteristics of people and countries in which they live (country-level), and cross-level interactions.

To test this model we need data at both the individual- and the country-level that vary over time and across space. The Harmonization Project sets out to create comparable measurements of political protest, social values, and demographics via ex-post harmonization of variables from international survey projects and append them with macro-level variables from external sources such as the World Bank, OSCE, UN agencies, Transparency International, and others.

Table 1 Selected International Survey Projects.

Abbrev.	Survey Project	Time span	Waves	Files	Data Sets	Cases
			Numbers			
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 in Central & Eastern Europe	1990-2001	2	1	27	28926
CNEP	Comparative National Elections Project	2004-2006	1	8	9	13978
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	6	6682
PA8NS	Political Action - An Eight Nation Study	1973-1976	1	1	8	12588
PPE7N	Political Participation and Equality in Seven Nations	1966-1971	1	7	7	16522
VPCPCE	Values and Political Change in Postcommunist Europe	1993	1	5	6	5769
Total		1966 -2013	89	81	1726	2293337

Note: In this table EVS and WVS are joined in one row because they share one data file. The total number of cases refers to all cases in source data files.

We selected 22 well-known international survey projects – listed in Table 1 – that span almost 50 years (1966-2013) and a total of 142 countries or territories.¹ In all survey projects, the units of observations are individuals. We took into account only projects designed primarily for academic use and with coverage of at least three countries. The data from selected projects are in the public domain, either in social science data archives or projects' own webpages that are open to scholars. Documentation of these projects is in English. Surveys contain political (e.g. protest), demographic (e.g. gender and age) and social stratification (e.g. education) items, but vary somewhat in their content and form.

From the selected projects, we pooled 81 data files, with 89 waves, into a relational database. It is a database containing 1726 national samples for which interviews were conducted in all waves (project wave country). All these surveys cover a total of almost 2.3 million respondents.² The platform for data files of national surveys is organized such that in the future any variable could be extracted and moved to the virtual integrated dataset (see Powalko 2014, and in this Newsletter).

We identified relevant original (source) variables that appear in at least five of the survey waves. Using various data processing procedures we produce, in the database, common (target) variables according to a unified measurement scheme. This scheme is well grounded in the past important discussions on ex-post harmonization (Gunther 2003; Minkel 2004; Ehling, Rendtel, et al. 2006; Granda and Blasczyk 2010; Granda, Wolf and Hadron 2010).

We select two types of source variables for harmonization: technical variables, provided by survey administrators, and variables of substantive interest. The list of variables is not closed, thanks to the flexible set-up of the programming environment we are using.

The Harmonization Project is work in progress. As it unfolds, it prompts us to reconsider how existing survey data can best be used in the harmonization framework by including controls of various quality aspects of existing surveys and harmonization procedures. We construct quality controls of the general survey documentation, the specific data description, and original data in the computer files. In addition, we apply quality control to specific harmonization procedures that could influence validity and reliability of the target variables. We suggest that quality-control variables for each of these aspects be included in substantive analyses (see Slomczynski and Tomescu-Dubrow in this Newsletter). Their relevance has to be empirically assessed.

NOTES

- 1 We refer to the selected projects as well-known on the basis of publication records and the impact that they have on the social science disciplines. For practical reasons, we stopped adding new data in the second quarter of 2014.
- 2 Because of the thematic coverage criterion, we include only survey waves that contain relevant questions on protest behavior and/or democratic values; thus, not all waves of ISSP, EB and CNEP are in our data.

REFERENCES

- Benson, Michelle and Thomas R. Rochon. 2004. "Interpersonal Trust and the Magnitude of Protest: A Micro and Macro Level Approach." *Comparative Political Studies* 3 (4): 435-457.
- Dalton, Russel J., Alix van Sickle, and Steven Weldon. 2009. "The Individual-Institutional Nexus of Protest Behaviour." *British Journal of Political Science* 40(1): 51-73.
- Dubrow, Joshua Kjerulf., Kazimierz M. Slomczynski, and Irina Tomescu-Dubrow. 2008. "Effects of Democracy and Inequality on Soft Political-Protest in Europe: Exploring the European Social Survey Data." *International Journal of Sociology* 38(3): 36 – 51.
- Dubrow, Joshua Kjerulf and Irina Tomescu-Dubrow. 2014. "A History of Cross-national Survey Data Harmonization Projects in the Social Sciences: Emergence of an Interdisciplinary Methodological Field." CONSIRT Working Paper Series 1 (CONSIRT Lab: Methodology of Survey Data Harmonization) at consirt.osu.edu.
- Ehling, Manfred and Ulrich Rendtel, et al. 2006. Synopsis. Research Results of Chintex - Summary and Conclusions: www.destatis.de/DE/Methoden/Methodenpapiere/Chintex/ResearchResults/Downloads/Synopsis.html [Last accessed November 28, 2014].
- Gelman, Andrew and Jennifer Hill. 2007. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. New York: Cambridge University Press.
- Granda Peter and Emily Blasczyk. 2010 *Cross-Cultural Survey Guidelines*. XIII. Data Harmonization. <http://ccsg.isr.umich.edu/harmonization.cfm> [Last accessed November 24, 2014].
- Granda Peter, Christof Wolf, and Reto Hadorn. 2010. "Harmonizing Survey Data." Pp. 315-332 in Janet A. Harkness, Michael Braun, Brad Edwards, Timothy P. Johnson, Lars Lyberg, Peter Ph. Mohler, Beth-Ellen Pennell, and Tom W. Smith (eds.), *Methods in Multinational, Multicultural and Multiregional Contexts*. Hoboken, NJ: John Wiley & Sons.
- Gunther, Ronald. 2003. "Report on Compiled Information of the Change from Input Harmonization to Ex-post Harmonization in National Samples of the European Community Household Panel – Implications on Data Quality." Wiesbaden: Statistisches Bundesamt - Working Paper 19.
- Kriesi, Hanspeter. 2004. "Political Context and Opportunity." Pp. 67-90 in David A. Snow, Sarah A. Soule and Hanspeter Kriesi (eds.) *Blackwell Companion to Social Movements*. Malden: Blackwell.

- Marien, Sofie, Marc Hooghe, and Ellen Quintelier. 2010. "Inequalities in Non-Institutionalized Forms of Political Participation: A Multi-level Analysis of 25 Countries." *Political Studies* 58(1): 187-213.
- Minkel, Hartmut. 2004. "Report on Data Conversion Methodology." CHINTEX Working Paper 20.
- Powałko, Przemek. 2014. "Working with Big Data: Experiences with the Cross-National Survey Data Harmonization Project." CONSIRT Working Papers Series (CONSIRT Labs: Methodology of Survey Data Harmonization) at consirt.osu.edu.
- Preacher, Kristopher J., Zhen Zhang, and Michael J. Zyphur. 2010. "Alternative Methods for Assessing Mediation in Multilevel Data: The Advantages of Multilevel SEM". *Structural Equation Modeling: A Multidisciplinary Journal* 18(2): 161-182.
- Schoene, Matthew and Marta Kołczyńska. 2014. "Survey Data Harmonization and the Quality of Data Documentation in Cross-National Surveys." CONSIRT Working Papers Series (CONSIRT Labs: Methodology of Survey Data Harmonization) at consirt.osu.edu.
- Vrablikova, Katerina. 2013. "How Context Matters? Mobilization, Political Opportunity Structure and Nonelection Political Participation in Old and New Democracies." *Comparative Political Studies* 47 (5), on line first.

Irina Tomescu-Dubrow is Associate Professor at the Institute of Philosophy and Sociology, Polish Academy of Sciences. She is Program Manager at Cross National Studies: Interdisciplinary Research and Teaching Program (CONSIRT), of the Polish Academy of Sciences and The Ohio State University.

Kazimierz M. Slomczynski directs CONSIRT. He also directs the Polish Panel Study 1988-2013 (POLPAN), a unique panel survey on the transformation of the Polish social structure.

Working with Data in the Harmonization Project

by **Przemek Powalko**

Polish Academy of Sciences

In the Harmonization Project we gather and process data from 22 international survey programs (see Tomescu-Dubrow and Slomczynski, Table 1 in this Newsletter). Given the substantive orientation of the project, we select those waves (i.e. project year) that contain relevant items on political protest and democratic attitudes. We end up with a database of 81 files with 1726 project wave countries (i.e. national samples in all projects and in all waves); the data base contains a total of almost 2.3 million observations (respondents).

Managing the sheer amount of data, as well as the variety of data formats applied in so many surveys, is a challenging task on its own. Additional problems arise when one requests to have all data at hand for quick and easy use. Instead of statistical packages typically used in the social sciences, for processing, combining and harmonizing data we have built a custom platform based on the concept of relational database and programmable Unix-like environment.

A programmable platform offers a simple way of writing scripts for repeatable procedures, which make all tasks fully automated, controllable, and fast. We use free and/or open source software.

A programmable platform offers a simple way of writing scripts for repeatable procedures, which make all tasks fully automated, controllable, and fast. We ground our solution on free and/or open-source software. We employ a relational database that (a) allows us to store data in tables (segments of records arranged in rows and columns), (b) guarantees mechanisms of integrity and consistency of data, and (c) enables sophisticated means of manipulating data with a high-level language, SQL. For reading source data files and converting them to plain text files we use PSPP, a free replacement for a proprietary statistical package, SPSS. Intermediate text files are further being processed and subsequently loaded to MySQL, an open-source database, which satisfies demands posed by the amount of data – not only in terms of the number of cases but also in terms of the number of columns that correspond to variables in the source data files (in extreme case being as high as 4096 per table). For browsing and querying data in the database, we use HeidiSQL, a free SQL editor. All scripting is done in Cygwin, a free environment which provides integration to Windows resources and a convenient interface for developers. Scripts themselves are written in free scripting languages such as SQL, Perl, and Unix-like shells.

All that makes our technical solution inexpensive and fast. Automation of tasks (batch jobs) is one factor. Another one is the way relational databases handle data:

executing an SQL query leads to a series of small data obtainable from tables – accessed (and in fact opened) on demand, so that, at all times, only a fraction of data is read from disk and loaded into internal memory. Another advantage of MySQL that we extensively make use of is its ARCHIVE engine, a mode of storing data on the hard disk, that significantly compresses data in tables. This makes the data much smaller and makes disk reads much faster. All these features make our solution not only fast but also highly scalable: virtually any number of tables can be stored and queried in database without losing flexibility, manageability, and performance.

It is beyond the purpose of this article to cover the details of harmonization itself. The one remark to make is that implementation of harmonization rules can be done through a series of simple SQL statements. At the end of the whole process we have created a single table, a master table, which can be exported from the database and converted to any format read by statistical packages such as STATA and R. This final product we call a master file. The structure of the master file is flexible, and may depend on end-user's expectations and needs. We plan to have at least three types of variables: source variables preserved for reference, target variables resulting from the application of the harmonization rules, and control variables. We plan the master file to be freely available online in the project's web page.

We developed a custom solution for processing data coming from numerous survey programs that exploits free software, including relational database and integrated development environment, and that allows many tasks to be fully automated via batch scripting. The data are being manipulated inside the database. The master file containing the harmonized variables is created with all information needed for further substantive analysis. The skills required to create a custom solution may be perceived as a drawback and suggests that our approach as an experiment. However, the aforementioned advantages that we have personally experienced have led us to believe that the undertaking is worth continuing.

Przemek Powalko works at the Institute of Philosophy and Sociology, Polish Academy of Sciences, as a computer specialist. He is responsible for data management in the Harmonization Project.

Survey Data Recycling: Toward a Formalized Approach to Ex-Post Harmonization of International Projects

by **Kazimierz M. Slomczynski and Irina Tomescu-Dubrow**
Polish Academy of Sciences and CONSIRT

In solving the many methodological challenges that the Harmonization Project is raising, we recognize the need to unify three major strands of survey research methodology that, up to now, have separate scientific literatures – survey data quality, harmonization, and multi-level modeling. We plan to do this by developing the analytic framework of survey data recycling.

We introduce the notion of survey data recycling as a novel way of approaching existing surveys to broaden the scope of substantive and methodological knowledge they can yield. Data recycling empowers scientists to reprocess existing survey information in a way that minimizes the “messiness” of data built into total survey error and, simultaneously, provides comparable measurements; it expands the range of data in terms of time and space, allowing researchers to introduce macro-level characteristics.

Our simple assumption is that paying attention to data quality, standardizing variables to achieve comparability, as well as taking advantage of the hierarchical structure of the data, improves confidence in substantive results. This is the goal of survey data recycling.

Data recycling is complex. In the case of international projects containing surveys conducted on national samples, its core involves:

- (a) organizing existing materials pertaining to the surveys taken into account,
- (b) creating common survey documentation,
- (c) evaluating the quality of original materials,
- (d) harmonizing survey variables, and
- (e) evaluating the quality of harmonization process.

Formal procedures need to be developed around each of these steps.

A programmable platform offers a simple way of writing scripts for repeatable procedures, which make all tasks fully automated, controllable, and fast. We use free and/or open source software.

In our approach, target variables T – that is, variables of substantive interest created through the harmonization process – are a function of original variables in surveys, so-called source variables S . The form of the relationship between T and S , $T = f(S)$, must be determined by researchers and depends on the substantive

problem and the availability of source variables. The novelty of survey data recycling rests in introducing quality control variables for target variables in the harmonization process, and employing this information in statistical analysis. We use two types of control variables, Q and H, in linear manner:

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

where Q stands for Data Quality Controls of general survey documentation, specific data description, and original data in the computer files; and H stands for Harmonization Quality Controls of specific procedures that could influence validity and reliability of T.

If b_1 and/or $b_2 > 0$, some intervention is needed to correct for errors in T. A possible solution, which we plan to assess empirically in statistical analyses, is to partial out the effects of Q and H on the relationships of T with other substantive variables, X. This would be a procedure analogous to computing partial correlation of T and X, controlling for Q and H.

Some data quality controls are, or could be, defined on the level of national surveys or even entire international survey projects. Thus, the equation joining T with Q and H must include subscripts reflecting the hierarchical structure of the data. This leads us to multi-level modeling.

Considering the hierarchical structure of the harmonized survey data is also important from substantive point of view. Generally, researchers may be interested in matching harmonized survey data with various characteristics of countries, coming from such sources such as official statistics provided by international organizations or scientific publications, among others. Survey data recycling offers various options of combining the harmonized and quality-checked survey file with information from non-survey sources.

The next two articles in this Newsletter, inspired by experiences within the Harmonization Project, can be read through the prism of data recycling. They discuss shortcomings in the quality of source surveys with regard to (a) general survey documentation (Kołczyńska) and (b) consistency between data description and records on the computer file (Wysmulek, Oleksiyenko, and Vangeli). Lack of, or inadequate information in, documentation reduces user confidence in the data. Inconsistencies of the resources defining the meaning of variables and their values with records on the computer data file decrease interpretability of the data. In the analytic framework of data recycling, these problems will be accounted for via quality-control variables, to produce a full-value product for researchers to use.

The Importance of Data Documentation for Survey Data Harmonization

by **Marta Kołczyńska**

The Ohio State University and Polish Academy of Sciences

Data, according to the United Nations Statistical Commission, are “the physical representation of information in a manner suitable for communication, interpretation, or processing by human beings or by automatic means” (UNSC 2000: 6). In other words, for information to qualify as data, it needs to be usable. Usable survey data depends on the availability and the high-quality of documentation.

Survey documentation refers to information on when, where, how and by whom the study was conducted, including information on the type of the sampling, size of the sample, response rate, preparation of the questionnaire and other instruments, as well as pretesting, and fieldwork control. In the Internet age, this information should accompany the survey data set in the form of one or more documents electronically available for viewing and downloading.

The main goal of any statistical analysis using survey data is to draw inferences about the target population. The precondition is that the survey sample is representative for the population. Representativeness can be approached in different ways and met to different degrees. The researcher ultimately has to decide whether a given survey sample is sufficiently representative to solve their research problem. This decision requires knowledge about sampling, including the sampling scheme, the sampling frame and, if such is the case, details of stratified samples or other methods. For researchers, additional aspects of the survey process, such as response rates and control of fieldwork, are also important to review in order to assess survey data quality.

In the case of cross-national studies, it is also advisable to review the survey tools, typically questionnaires and the process of their creation, including what translation procedure was applied, and whether the questionnaires were pretested. Best practices for translation are debated in the field of survey methodology (see e.g. Harkness, Pennell, and Schoua-Glusberg 2004; Harkness, Villar, and Edwards 2010). However, the consensus is that high quality translation is a prerequisite for comparability of data collected in different linguistic and cultural contexts. Information on the translation procedure must be provided in the survey documentation for a given country.

Pretesting is not only a way of validating the translation to avoid information loss or changes in the meaning of the basic concepts; it is also a way to assess the degree to which the questionnaire meets the criteria of acculturation (i.e. to

what extent it fits to the mindset of potential respondents). If information about pretesting is lacking or inadequate, then, justifiably, researchers have lower confidence in the data.

Similarly, high quality surveys usually perform some kind of fieldwork control that typically consists of a personal visit or phone call to back-check the previously collected data. Regardless of the method, fieldwork control is generally beneficial because it improves interviewers' performance. Again, if there was no fieldwork control or information about it is not provided in survey documentation, researchers worry about the quality of that data.

Documentation – at least in the case of surveys – is an integral part of the data. Information about sampling, response rate, translation of the questionnaire, pretesting and fieldwork control cannot be found in the numerical data recorded in computer files, but it is important for interpretation of these data. In the case of comparative studies, variations in documentation quality within and across international projects should be recorded as survey-quality indicators.

To qualify as data, information needs to be usable. Usable survey data depends on the availability and the high-quality of documentation.

Working within the Harmonization Project makes this point clear. In searching through the documentation of the 22 international survey projects listed in Table 1 in this Newsletter, my colleagues and I have found wide variation in the standards of documentation accompanying each data set. At this point we created five variables describing data documentation of all 1726 national surveys: (1) response rate – whether this information is provided or not, (2) numerical value of response rate, if given, (3) indication of any efforts at controlling the quality of the questionnaire translation, (4) whether there is any indication of questionnaire pretesting, and (5) attempts of the fieldwork control (Schoene and Kołczyńska 2014). With the exception of numerical values of the response rate, all other variables are dummies (1 – yes, 0 – otherwise). The distribution of all these variables differentiates national surveys enough to claim that surveys from the selected international projects are of varying quality.

We aim to build documentation quality controls into statistical analyses of the Harmonization Project database, to check empirically the consequences of weak documentation standards in cross-national projects. In doing so, we hope to contribute to the discussions about how to increase confidence in extant cross-national survey data.

REFERENCES

Economic Commission for Europe of the United Nations (UNECE). 2000. "Terminology on Statistical Metadata", Conference of European Statisticians Statistical Standards and Studies, No. 53, Geneva.

- Harkness, Janet A., Beth-Ellen Pennell, and Alisú Schoua-Glusberg. 2004. "Questionnaire Translation and Assessment." Pp. 453-472 in Stanley Presser, Jennifer Rothgeb, Mick Couper, Judith Lessler, Elizabeth Martin, Jean Martin, and Eleanor Singer (eds.), *Methods for Testing and Evaluating Survey Questionnaires*. Hoboken, NJ: John Wiley & Sons.
- Harkness, Janet A., Ana Villar, and Brad Edwards. 2010. "Translation, Adaptation and Design." Pp. 117-139 in Janet A. Harkness, Michael Braun, Brad Edwards, Timothy P. Johnson, Lars Lyberg, Peter Ph. Mohler, Beth-Ellen Pennell, and Tom W. Smith (eds.), *Methods in Multinational, Multicultural and Multiregional Contexts*. Hoboken, NJ: John Wiley & Sons.
- Schoene, Matthew and Marta Kołczyńska. 2014. "Survey Data Harmonization and the Quality of Data Documentation in Cross-National Surveys." CONSIRT Working Papers Series (CONSIRT Labs: Methodology of Survey Data Harmonization) at consirt.osu.edu.
- UNSC – United Nations Statistical Commission and Economic Commission for Europe. 2000. *Terminology of Statistical Data*. Geneva: United Nations.

Marta Kołczyńska is a PhD student at the Department of Sociology, The Ohio State University, and a research assistant in the Harmonization Project.

Processing Errors in Cross-national Surveys: Insights from the Harmonization Project

by Ilona Wysmulek, Olena Oleksiyenko, and Anastas Vangeli

Polish Academy of Sciences

A taxonomy of survey errors – that is, a comprehensive checklist of all possible errors that could affect the quality of the survey – is elaborated upon within the Total Survey Error (TSE) approach (e.g., Weisberg 2005). Another approach to assess survey quality is Total Quality Management (TQM) (e.g., Morgenstein and Marker 1997), which emphasizes that all stages of survey production singularly, and in conjunction with each other, have a direct influence on the overall quality of the end-product (for a comparison of TQM and TSE, see Loosveldt, Carton, and Billiet 2004: 66).

Of all the elements of the preparation and administration of survey fieldwork, relatively little attention has been paid to “processing error”. Processing error refers to the transparency and consistency of documentation. At this stage of research, both systematic and random errors could occur that might (similarly to measurement errors, sampling errors and nonresponse) undermine the overall reliability of the survey. Being first introduced by Deming (1944), processing errors are presently included in almost all modern taxonomies of errors, among others, in categorization of Anderson et al (1979), Groves (1989), and Biemer and Lyberg (2003); they are also called “compiling errors” (Hansen, Hurwitz, and Madow 1953), and “survey administrative issues” (Weisberg 2005).

Different types of errors caused by various mistakes following data collection, considered within the Total Survey Error (TSE) approach, refer to “coding, editing, imputation, and other data processing activities that follow the data collection phase” (Groves 1989: 12). However, in practice this error component in TSE “... is too rarely included in models of survey error” (Groves and Lyberg 2010: 869). The Harmonization Project deals with processing errors explicitly, by focusing on the quality of the correspondence between the documentation and the data in the computer files.

Of all the elements of the preparation and administration of survey fieldwork, relatively little attention has been paid to processing error.

We check the consistency between the survey documentation and survey data in the computer files of 22 international survey projects (for the list, see Table 1 in Tomescu-Dubrow and Slomczynski in this Newsletter). We analyze the errors that occur in the case of selected individual variables. Information on any given variable is extracted from codebooks, questionnaires, SPSS dictionaries, and

the data contained in computer files. From each source of documentation, we recorded variable name, question number, exact question formulation, variable label, and value labels, and we perform inter-source comparison. We recorded all of the discrepancies that appeared between sources and created a typology of possible errors that can occur between documentation and data in computer files. Examples of quality-control variables that we constructed include: (1) variable value discrepancy, (2) contradictory value labeling, (3) lack of value labels, (4) misleading variable label, and (5) insufficient information about variable meaning.

From our personal experience, checking for processing errors is an enormous time investment, and it requires, in equal measure, tenacity, creativity and careful attention to detail. It cannot (as of yet) be computer automated: it can only be done with human beings' unique power of discernment. In the context of data harmonization, this process has proven to be worthwhile, as checking for processing errors is essential to get a sound understanding of data quality.

REFERENCES

- Anderson, Ronald, Judith Kasper, Martin Frankel, and Associates. 1979. *Total Survey Error: Applications to Improve Health Surveys*. San Francisco, CA: Jossey-Bass.
- Biemer, Paul P. and Lars E. Lyberg. 2003. *Introduction to Survey Quality*. Hoboken, NJ: John Wiley & Sons.
- Deming, W. Edward. 1944. "On Errors in Surveys." *American Sociological Review* 9(4): 359–369.
- Groves, Robert M. 1989. *Survey Errors and Survey Costs*. New York: John Wiley & Sons.
- Groves, Robert M. and Lars E. Lyberg. 2010. "Total Survey Error: Past, Present, and Future." *Public Opinion Quarterly* 74 (5): 849-879.
- Hansen, Morris H., William N. Hurwitz, and William G. Madow. 1953. *Sample Survey Methods and Theory*. New York: Wiley.
- Loosveldt, Geert., Ann Carton, and Jaak Billiet. 2004. "Assessment of Survey Data Quality: A Pragmatic Approach Focused on Interviewer Tasks." *International Journal of Market Research* 46 (1): 65-82.
- Morgenstein, David and David A. Marker. 1997. "Continuous Quality Improvement in Statistical Agencies." Pp. 475-500 in Lars Lyberg, Paul P. Biemer and Martin Collins (eds.), *Survey Measurement and Process Quality*. New York: Wiley.
- Smith, Tom W. 2011. "Refining the Total Survey Error Perspective." *International Journal of Public Opinion Research* 23(4): 464-484.
- Weisberg, Herbert F. 2005. *The Total Survey Error Approach: A Guide to the New Science of Survey Research*. Chicago: University Of Chicago Press.

Ilona Wyszumłek, Olena Oleksiyenko and Anastas Vangeli are PhD students at the Graduate School for Social Research at the Polish Academy of Sciences. They are research assistants in the Harmonization Project.

Building a Community

by Joshua Kjerulf Dubrow and Irina Tomescu-Dubrow

CONSIRT

This is the second issue of Harmonization: Newsletter on Survey Data Harmonization in the Social Sciences. We continue to share news and communicate with the growing community of scholars, institutions and government agencies who work on harmonizing social survey data and other projects with similar focus.

This issue features articles from harmonization and data quality scholars from the USA and Europe. The first article is by Peter Granda, who writes that “Researchers spend extensive time and resources in creating harmonized datasets; they should take the few extra steps necessary to make certain their hard work is preserved for future users.” To achieve this goal, researchers should use available tools to conserve the codes used for the transformations of the original source variable into the target variables, together with the full documentation of the questionnaire items and datasets. As Granda points out, the metadata standard of the Data Documentation Initiative (DDI) is a right way of safeguarding harmonization process.

Next are two articles on harmonization of specific variables: Verena Ortmanns and Silke L. Schneider write about educational attainment comparability, referring to the project on Computer-Assisted Measurement and Coding of Educational Qualifications in Surveys. Wiebke Breustedt’s paper is ex-ante output harmonization of trust in institutions across regional barometers.

The last two articles propose some solutions to problems Peter Granda highlights. Kristi Winters announces the CharmStats, a free and open-source harmonization software product developed at GESIS - Leibniz Institute for the Social Sciences, which allows users to document variable harmonization. Wismulek and her co-authors present a template for target variable reports, with the purpose of ensuring the transparency and replicability of harmonization procedures. Reports refer to rules of transforming source variables into target variables, including construction of control variables.

As always, we invite you to send us your research on survey data harmonization to share with the community.

Archiving and Preserving the Relationships between Harmonized Survey Datasets and Their Sources

by Peter Granda

University of Michigan – ICPSR

Data harmonization opens new research possibilities both for producers of new datasets and for the social scientists who will use them. To deal with the complex interactions between harmonized files and the original sources from which they derived, more and more producers create sophisticated web portals that provide users with multiple paths and strategies to engage this material. In addition to download capabilities for data and documentation, these portals often include online analysis, the ability to compare questions and responses as source variables are transformed into target variables, an assessment of the quality of the harmonization process, and immediate access to training resources and other aids to understand how to analyze the data properly. Because producers can provide the necessary links and documentation to describe the entire process, in many ways, a web dissemination ‘environment’ is ideally suited to the presentation of all the relationships that exist when harmonized files are created.

But web portals do not last forever and they are an inadequate setting for the archival preservation of harmonization materials. What should happen to safeguard harmonization projects in the long-term?

The key is to preserve all of the contents and, most importantly, the associations between the contents. Users must be able to see in as much detail as possible how source variables became target variables. Among the principal elements that require preservation are: the code used for the transformations, particularly if it is recorded in some standard statistical package such as SAS, SPSS, or Stata; the original source variable names; and the original source questions and datasets. It is preferable to preserve this information in the lowest common denominator possible: raw text (ASCII) or a familiar software such as EXCEL that might be easily converted to text if necessary.

Below is an example where such information is preserved in a spreadsheet for a harmonization project that involved ten cross-sectional surveys on the subject of fertility in the United States from 1955–2002:

Harmonized Variable Name	Variable Group	Source Study Name	Source Variable	SOURCE_QUESTION
IFSS_ABORT12	Family planning and medical services	National Survey of Family Growth, Cycle VI, 2002	ABORT12	FA-3b
IFSS_ABORT12	Family planning and medical services	National Survey of Family Growth, Cycle V, 1995	ABORT12	FB-2
IFSS_ACHIEVE	Sociodemographic characteristics	National Survey of Family Growth, Cycle VI, 2002	ACHIEVE	IH-12
IFSS_ACHIEVE	Sociodemographic characteristics	National Survey of Family Growth, Cycle V, 1995	ACHIEVE	ID-6
IFSS_ACHIEVE	Sociodemographic characteristics	National Fertility Survey, 1970	ATT_MOMHM	CM327/PM218
IFSS_ADEXP5YR	Birth desires and intentions	National Survey of Family Growth, Cycle II, 1976: Couple File	ADEXP5YR	D-39
IFSS_ADEXP5YR	Birth desires and intentions	National Survey of Family Growth, Cycle III, 1982	D28_35	D-28/D-35
IFSS_ADEXP5YR	Birth desires and intentions	Growth of American Families, 1960	EXP_MAX5YR	27a
IFSS_ADEXP5YR	Birth desires and intentions	Growth of American Families, 1960	EXP_MIN5YR	27a

Another and potentially more powerful option for preserving all of the relationships between source and target variables is to store the information in XML using a metadata standard such as the Data Documentation Initiative (DDI), which pertains specifically to social science survey data. The markup of the content makes it machine-readable; the textual basis of XML makes it ideal as a preservation medium; and each XML element defines a separate characteristic of each variable.

As an example, the following few lines of XML define the value of category 5 of the question: SC8_1 How would you rate your overall physical health - excellent, very good, good, fair, or poor? The meaning of category 5 for this question is “poor”. The question was asked in four other languages besides English: Spanish, Tagalog, Vietnamese, and Chinese. The words for “poor” in each language appear followed by the unweighted frequency (188) and percent (4.0) of respondents who answered this question in one of the source data files.

```
<catgry missing="N" source="producer" excl="true">
  <catValu source="producer">5</catValu>
  <labl level="category" xml-lang="en" source="producer">POOR</labl>
```

```

<labl level="category" xml-lang="es" source="producer">POBRE</labl>
<labl level="category" xml-lang="tl" source="producer">MAHINA</labl>
<labl level="category" xml-lang="vi" source="producer">kém</labl>
<labl level="category" xml-lang="zh" source="producer">不好</labl>
<catStat type="freq" source="producer" wgt="not-wtgd"> 188 </catStat>
<catStat type="percent" source="producer" wgt="not-wtgd"> 4.0 </catStat>
</catgry>

```

The next XML excerpt identifies this same variable (V00233), now as a target variable, as it appears in the harmonized dataset (CPES = Collaborative Psychiatric Epidemiology Surveys) and its comparable variables from two source files (NLAAS = National Latino and Asian American Study [SC8_1] and NSAL = National Survey of American Life) [C8]).

```

<varFormat type="numeric" source="producer" schema="ISO"
category="other">ASCII</varFormat>
<notes type="harmonized-variable" subject="CPES" source="producer">
http://www.icpsr.umich.edu/icpsrweb/ICPSR/ssvd/studies/20240/datasets/0001/variables/V00233</notes>

```

```

<notes type="comparable-variable" subject="NLAAS" source="producer">
http://www.icpsr.umich.edu/icpsrweb/ICPSR/ssvd/studies/20240/datasets/0004/variables/SC8\_1</notes>

```

```

<notes type="comparable-variable" subject="NSAL" source="producer">
http://www.icpsr.umich.edu/icpsrweb/ICPSR/ssvd/studies/20240/datasets/0003/variables/C8</notes>

```

Researchers spend extensive time and resources in creating harmonized datasets; they should take the few extra steps necessary to make certain their hard work is preserved for future users.

Peter Granda is Associate Director of the Inter-university Consortium for Political and Social Research (ICPSR). He also directs both the General Archive and the Health and Medical Care Archive, a topical archive supported by the Robert Wood Johnson Foundation. He has a long association with the cultures of South Asia, where he spent several years of study in the southern part of the Indian subcontinent.

Harmonization of Educational Attainment Variables in Cross-national Surveys: The CAMCES-Project

by Verena Ortmanns and Silke L. Schneider

Educational attainment is a widely used variable in survey research. However, its precise measurement varies over time, between countries, and across surveys. Output harmonization procedures are meant to mitigate this incomparability. In recent research (Ortmanns and Schneider, 2015), we examined education harmonization outcomes in the Eurobarometer (EB), the European Social Survey (ESS), the European Values Study, and the International Social Survey Programme (ISSP). We found discrepancies in the distributions of this variable. Those discrepancies can most likely be explained through inconsistent coding. We had to conclude that the harmonization of educational attainment data in the cross-national context is still a challenge. As a possible solution, we would like to briefly present ongoing work on new survey tools and information resources that may, in the future, provide a way to address the underlying problems.

We begin with more background information on harmonization. Educational systems differ substantially across countries and some educational qualifications cannot be translated. Cross-national surveys measure educational attainment comparably using an approach called ex-ante output harmonization. This process is designed to ensure that the development of data collection instruments (which in the case of education involve country-specific response categories) and the coding of the resulting variable into an international coding framework (i.e. standard classification) result in comparable data. There is an order to this process. The international coding framework or standard classification, as well as the relationship between country-specific categories and international codes, have to be specified before finalizing the data collection instruments. Most comparative surveys these days use the International Standard Classification of Education (ISCED).

In principle, pooling data from different cross-national surveys and comparing variables that were harmonized using the same standard classification or coding framework should be possible. In order to do this, the variables have to be coded consistently. However, since ISCED mappings are sometimes contested, different coders may choose different ISCED codes for the same qualification, producing “deliberate misclassifications”. Coding inconsistencies such as these can only be detected if the harmonization process is transparently documented. The quality criteria for consistent coding and transparency, amongst others, were formally laid down in the “European Statistics Code of Practice” by Eurostat and European Statistical System Committee (2011). However, transparency was not always

provided for official surveys such as the European Union Labour Force Survey (EU-LFS) before 2014, the European Union Statistics on Income and Living Conditions (EU-SILC), or the Programme for the International Assessment of Adult Competencies (PIAAC). For those surveys, documentation of the harmonization approach, especially the correspondence between country-specific response categories and ISCED-categories, is not publicly available. In addition, the country-specific variables are not included in the data sets of the surveys. Therefore it is not possible to check these data for inconsistent coding and to compare them with other surveys.

In the project "Computer-Assisted Measurement and Coding of Educational Qualifications in Surveys" (CAMCES), we are thus currently working on new open-source survey tools to facilitate the reduction of such coding inconsistencies, and more generally improve the quality of educational attainment data. The tools consist of a question module, an international qualifications database, a software interface, and standard harmonization routines. Together, they enable accurate and detailed reporting and cross-nationally comparable coding of the highest educational qualification obtained. For example, migrants can more easily and accurately indicate foreign educational qualifications than with the "show card procedure" described above. The database will also be useful for ex-post harmonization of education variables. The tools will be published towards the end of 2016. They can be implemented in computer-assisted surveys (CAPI and CAWI) and will initially cover all European countries. The scope of countries will be increased and fields of education be added within a follow-up project, which is part of the project "Synergies for Europe's Research Infrastructures in the Social Sciences" (SERISS) coordinated by City University, London, funded through Horizon2020.

CAMCES website: <http://www.gesis.org/en/research/external-funding-projects/projektuebersicht-drittmittel/camces/>

SERISS website: <http://seriss.eu/>

REFERENCES

- Eurostat. & European Statistical System Committee. (2011). European Statistics Code of Practice for the National and Community Statistical Authorities. Retrieved from <http://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/KS-32-11-955>
- Ortmanns, V., & Schneider, S. L. (2015). Harmonization Still Failing? Inconsistency of Education Variables in Cross-National Public Opinion Surveys. *International Journal of Public Opinion Research*, online first publication. doi: 10.1093/ijpor/edv025.

Verena Ortmanns is a junior researcher at GESIS – Leibniz Institute for the Social Sciences in Mannheim, Germany, and she is working in the CAMCES-project. She is a PhD student at the Justus Liebig University Giessen.

Silke L. Schneider is a senior researcher and consultant at GESIS – Leibniz Institute for the Social Sciences in Mannheim, Germany, and she is the head of the CAMCES-project. She focuses on research and consulting on the measurement of socio-demographic variables in survey research.

The Barometer Surveys: Insights into the Quality of the Harmonized Political Trust Items

by Wiebke Breustedt

In the social sciences, empirical research on political trust is commonplace (Newton 2015, 19). Yet, there are relatively few cross-national comparative analyses of political trust in countries outside Europe and the U.S. This is despite significant developments in the availability of global public opinion surveys over the past decades (Norris 2009, 522). The World Values Survey (WVS) is one of the most extensive cross-national survey projects (Heath et al. 2005, 302). Social scientists in the WVS network strive to make the WVS data comparable by means of a common questionnaire, i.e. by harmonizing them *ex ante* to a certain extent (WVS 2015).¹

Recently, the Global Barometer surveys (GBS) - the Afrobarometer, the Asian Barometer, the Arab Barometer and the Latinobarómetro - provide a complementary perspective to the WVS. They address economic and political attitudes in detail, including political trust (GBS 2015a). In terms of method, these regional barometers have not (yet) developed a common *ex ante* harmonization strategy.² Nevertheless, they represent promising candidates for *ex post* harmonization. That is to say, while their surveys were not originally designed to be comparable, comparability may be established to a certain extent after the data collection process (Information Society Technologies and CHINTEX 1999, 2). For example, the GBS network has published ‘GBS module 1’, the first *ex post* harmonized data set based on the regional barometer surveys (GBS 2015c).

Prior to using harmonized comparative data sets, researchers should carefully consider their quality. To determine the use of the harmonized data for valid comparisons across countries, survey quality assessments are important. I seek to contribute to the endeavor of assessing the survey quality of the GBS module 1 by providing preliminary insights into the quality of the harmonized political trust items.

There are three factors that determine the use of harmonized survey data for comparative research (Survey Research Center 2010, II-4-5, XIII-9-13): First is the quality of the original survey data; Second is the quality of the harmonized data; And third is the quality of the harmonization process. These factors can be considered with regard to the survey as a whole, as well as to the individual items (Survey Research Center 2010, XIII-8). As part of the Harmonia project, Schoene and Kołczyńska (2014) have addressed the first factor. Given the focus of this contribution on political trust, I will address the quality of the harmonized data and

the quality of the harmonization process as they pertain to the political trust items included in the GBS 1 module.

The quality of the harmonized political trust items can be assessed in terms of comparability and completeness (Granda et al. 2010, 322). The comparability of the items from the different Global Barometer surveys varies depending on the comparability criterion (Kiecolt and Nathan 1985, 56-62; Granda et al. 2010, 322-325). The political trust items are generally administered as a battery of items, but the sequence differs by international survey project. In the Latinobarómetro, for example, the items on trust in the president and trust in the government are asked separately from the questions on trust in other political objects. The question wording in the regional barometer surveys also varies. The number of response categories in the rating scales is comparable (4-point scale). The rating scales in all of the surveys are fully anchored. The anchors of the rating scale categories and the categories outside the rating scale (e.g. “don’t know,” “can’t choose,” and the like) are the same in the Asian Barometer and the Arab Barometer.³ The categories in the Latinobarómetro and the Afrobarometer are similar. The order of the anchors is the same in the Asian Barometer, the Latinobarómetro and the Arab Barometer questionnaires, the lowest category indicating the highest level of trust and the highest category indicating the lowest level of trust.⁴

“Completeness” is a quality criterion of the harmonized political trust items. It concerns the “degree to which the original information is preserved in the harmonized data” (Granda et al. 2010, 322). Since the number of response categories in the rating scale was not changed, the information is preserved in the GBS module 1. The module does not include all of the countries available such as Yemen and Cambodia, for example.

The quality of the harmonization process as it pertains to the individual items can be assessed in terms of consistency, i.e. the extent to which the data in the GBS module 1 are consistent with other harmonization efforts.⁵ The data are generally consistent in terms of frequencies.

There are three respects in which GBS 1 module is inconsistent. For one, there are deviations in the case of Lebanon with regard to all of the political trust items, and in the case of Palestine with regard to the item measuring trust in the prime minister. Since the GBS network has not published any documentation of the harmonization process, it is not possible to discern the reason for these discrepancies. In addition, it includes data on trust in the electoral commission for Singapore. The Asian Barometer survey file notes, however, that the item in Singapore asked respondents about trust in international TV. This note is not included in the GBS module 1. Third, the original Afrobarometer survey does not include a survey item asking about trust in political parties in general but rather about trust in the ruling parties and the opposition parties. The GBS module 1

includes data on trust in political parties for the African countries but does not explain how the data were calculated.

CONCLUSION

There are several quality issues with regard to consistency and completeness in the GBS 1 module. Nevertheless, that the political trust items are comparable in several respects makes them promising candidates for harmonization. There are four reasons why researchers should harmonize the regional barometer data themselves.

1. The harmonization process underlying the GBS module 1 is not transparent as the GBS network has not published any documentation in this respect.
2. The GBS module 1 only includes a small percentage of the items available in the regional barometer surveys. Depending on the case selection and the topic of interest, researchers can make better use of the data available by considering the original surveys and harmonizing the data themselves.
3. Harmonizing the data oneself increases the awareness of the differences between the survey items outlined above, thereby instilling the necessary caution when interpreting the data.
4. To date, there are many more waves of the regional barometer surveys available than are included in the GBS module 1.

Overall, harmonizing the regional barometer surveys is a worthwhile endeavor as it permits extended empirical studies of political trust and other political attitudes in terms of countries, time, causes and effects.

NOTES

- 1 Ex ante harmonization can involve a whole range of additional strategies (see Grais 1999, 65 and van Deth 2009, 88).
- 2 While there is no detailed information available on the GBS website, apparently, the members are currently establishing a standard approach (GBS 2015b). This is further corroborated by the fact that the Asian Barometer country teams have “to comply with the research protocols developed and established by the Global Barometer network” (Asian Barometer 2015). In addition, the GBS network has developed a “global question module on attitudes toward democracy” (Afrobarometer 2015).
- 3 Except for the fact that the Asian Barometer includes the option ‘do not understand the question’.
- 4 A detailed analysis of the comparability of the question wording and response categories should consider the country-specific questionnaires.
- 5 Since the number of response categories in the harmonized data file does not differ from the number of categories in the original files, in this case, consistency can be assessed by comparing the GBS module 1 with the original merged regional barometer files.

REFERENCES

- Afrobarometer 2015: Our History. <http://afrobarometer.org/about/our-history> [Last accessed September 10, 2015].
- Asian Barometer. 2015. Surveys. Survey Methods. <http://www.asianbarometer.org/newenglish/surveys/SurveyMethods.htm> [Last accessed September 10, 2015].
- Global Barometer Surveys. 2015a. Strategy. <http://www.globalbarometer.net/page/strategy> [Last accessed September 10, 2015].
- Global Barometer Surveys. 2015b. Background. <http://www.globalbarometer.net/page/background> [Last accessed September 10, 2015].
- Global Barometer Surveys. 2015c. Data. <http://www.globalbarometer.net/data/data> [Last accessed September 10, 2015].
- Grais, Bernard. 1999. "Statistical Harmonisation and Quality. The Case of Social Statistics." Pp. 51-114 in European Commission (ed.), *The Future of European Social Statistics. Harmonisation of Social Statistics and Quality. The Fourth Mondorf Seminar. 26 and 27 March 1998*. Luxembourg: Office for Official Publications of the European Communities.
- Granda, Peter, Christof Wolf, and Reto Hadorn. 2010. "Harmonizing Survey Data." Pp. 315-332 in Janet A. Harkness, Michael Braun, Brad Edwards, Timothy P. Johnson, Lars Lyberg, Peter Ph. Mohler, Beth-Ellen Pennell and Tom W. Smith (eds.), *Survey Methods in Multinational, Multiregional, and Multicultural Contexts*. Hoboken, NJ: John Wiley & Sons.
- Heath, Anthony, Stephen Fisher and Shawna Smith. 2005. "The Globalization of Public Opinion Research." *Annual Review of Political Science* 8: 297 - 333.
- Information Society Technologies and CHINTEX. 1999. Synopsis. The Change from Input-Harmonisation to Ex-Post Harmonisation in National Samples of the European Community Household Panel. Implications on Data Quality. https://www.destatis.de/DE/Methoden/Methodenpapiere/Chintex/ResearchResults/Downloads/Synopsis.pdf?__blob=publicationFile [Last accessed September 10, 2015].
- Kiecolt, K. Jill and Laura E. Nathan. 1985. *Secondary Analysis of Survey Data*. Beverly Hills: Sage.
- Newton, Kenneth. 2015. "Trust and Political Disenchantment. An Overview." Pp. 19-30 in Christina Eder, Ingvill C. Mochmann and Markus Quandt (eds.), *Political Trust and Disenchantment with Politics. International Perspectives*. Leiden: Brill.
- Norris, Pippa. 2009. "The Globalization of Comparative Public Opinion Research." Pp. 522-540 in Todd Landman and Neil Robinson (eds.), *The SAGE Handbook of Comparative Politics*. Los Angeles: SAGE.
- Schoene, Matthew and Marta Kolczyńska. 2014. "Survey Data Harmonization and the Quality of Data Documentation in Cross-National Surveys." *CONSIRT Working Paper October 2014*. <http://consirt.osu.edu/wp-content/uploads/2014/11/CONSIRT-Working-Papers-Series-3-Schoene-and-Kolczynska.pdf> [Last accessed September 10, 2015].
- Survey Research Center. 2010. Guidelines for Best Practice in Cross-Cultural Surveys. Ann Arbor, MI: Survey Research Center, Institute for Social Research, University of Michigan. <http://www.ccsr.isr.umich.edu/> [Last accessed September 10, 2015].
- World Values Survey. 2015. Who We Are. <http://www.worldvaluessurvey.org/WVSContents.jsp> [Last accessed September 10, 2015].

Wiebke Breustedt is Ph.D. student at the University of Cologne (Germany). She worked as research assistant at the Chair for Comparative Politics at the University of Duisburg-Essen (Germany) and as lecturer at the Chair for Comparative Politics at the University of Cologne. Her research and publications address questions pertaining to political attitudes (with a particular focus on political trust), social science research methods, political psychology and empirical democratic theory.

The CharmStats Program for Survey Data Harmonization

by **Kristi Winters**

GESIS – Leibniz Institute for the Social Sciences

Statistical analyses oftentimes require data harmonization; however there are no academic standards for harmonization documentation to facilitate transparency and replication. QuickCharmStats 1.1 for PC is the digital solution to the problem of documenting variable harmonization. The CharmStats workflow collates metadata documentation, meets the scientific standards of transparency and replication, and encourages researchers to publish their harmonization work. CharmStats is the name of a line of free and open-source harmonization software products. Developed at GESIS - Leibniz Institute for the Social Sciences, QCS is available for download on the CharmStats website. The software is based on logical workflow that allows users to import and work with all the metadata necessary to document variable harmonization. Its features include automated harmonization syntax generation for SPSS or Stata, creating reports that are publishable as either html or PDFs, and a graph generator that displays source and target response mapping for visual inspection. Users can enter extended notes on coding decisions as, for example, how missing cases were handled in the survey of a particular nation. Finally, QCS produces comprehensive digital harmonization projects that can be submitted for publication with GESIS. Users can enter in extended notes on coding decisions as, for example, how missing cases were handled in the survey of a particular nation.

QuickCharmStats is a Java® based desktop application. The installation package for QCS is available from the GESIS website. Users must provide a name and email address; after, a link for the download is sent to their email accounts. The QCS software, its code, the user manual and practice datasets come as part of the download. The software supports versions of Java 1.6 or higher. CharmStats products work by storing persistent information in a relational database. QCS works on a local database instance (localhost). We chose MySQL DBMS as the system to manage the database and its content. MySQL (Structured Query Language) is an open-source relational database management system (RDBMS) owned by Oracle Corporation. Users can download MySQL software for free if they do not have it.

After copying the QCS zip-file to the place of installation and unzipping, users connect QCS to the MySQL database by running a setup batch file in the windows command processor (cmd.exe) once. To start the application, double-click the CStatsApp jar-file symbol in the QuickCharmStats directory. The program is ready for use.

Until CharmStats, collecting the metadata to document variable harmonization work was too time consuming. QuickCharmStats organizes a researcher's work in a format pre-prepared for the peer-review process. If the harmonization documentation project is accepted for publication it will receive a permanent identifier that can be listed as a reference. Clicking the permanent identifier link will take the reader directly to the published harmonization documentation making their work available to all. Researchers can download the peer review submission report template from the GESIS CharmStats website. The completed report and the graph should be emailed to: charmstats@gesis.org. Once accepted, the submitted digital harmonization will be assigned a DOI and published on a GESIS website or deposited into an online harmonization library, free of cost. To learn more, please visit the CharmStats website.

Dr. Kristi Winters is Project Manager for CharmStats at GESIS. Her doctoral research investigated the construct validity of the man/woman variable as interpreted in the political science. She is the founder of the Qualitative Election Study of Britain, a study that is applying harmonization techniques to its longitudinal data collection.

Towards Standardization: Target Variable Report Template in the Harmonization Project

by Ilona Wysmulek, Olena Oleksiyenko, Przemek Powalko, Marcin W. Zieliński, and Kazimierz M. Słomczynski

We present a template for target variable reports, developed in the Data Harmonization project after a long search for ways to standardize the documentation of harmonization procedures, with the purpose of ensuring their transparency and replicability. This template, which we intend to further use in the Survey Data Recycling program, might be helpful to researchers involved in ex-post survey harmonization, as means to document the process of constructing target variables.

The structure of the target variable report is compliant with the newest Data Documentation Initiative standard (version 3.2) and focuses on data processing module (DDI, 2014). The report builds on experience of similar projects such as CHINTEX (Günther, 2003), SHARE (Phillips, Chien, Angrisani, Meijer and Lee, 2014), and CharmStats (Winters and Martin, 2015).

One of the crucial aims of the Data Harmonization project is to enable the transparency of harmonization procedures by providing clear information about (a) the decision-making process of transforming source variables into target variables, (b) the specific features of source variables, and (c) the quality of the analyzed surveys by means of control variables.

Target variable reports

Each target variable in the Data Harmonization project is accompanied by the following documents:

1) General target variable report – the document contains definition and operationalization of a target variable, enumeration of international survey projects involved (with waves, countries, and years), reference to documentation of source variables, and rules of transforming source variables into a target variable, including harmonization control variables.

2) Detailed target variable report – this document (in Excel format) contains question wording and response categories of source variables, target variable codes, and control variable codes.

3) Target variable syntax file – the document with a complete SQL syntax implementing the harmonization rules, i.e., the code transforming source variables into the target variable.

Source variable (S) – variable collected from an original survey for the ex-post harmonization.

Target variable (T) – variable created through the ex-post harmonization, which is a function of one or more source variable: $T = f(S)$.

Control variable (C) – variable controlling specific features of source variable(s) and survey's quality, i.e., potentially influencing validity and reliability of T.

Documenting the target variable in such a detailed manner allows for replicability of the harmonization process, and provides quality control of each stage of work with the variable. Control variables provide researchers with the opportunity to adjust a definition of the target variable to meet their particular goals and the flexibility to handle special cases.

General Target Variable Report – Template

Below we present an annotated template of the general target variable report extracted from the Rural/Urban Locality ($T_{RURALURB}$) report. For comprehensibility, we use blue color to indicate our clarifying comments.

TARGET VARIABLE REPORT – RURAL/URBAN LOCALITY

Prepared by: a person responsible for the report

Research team: list of people responsible for harmonization of the target variable

October 13, 2015

GENERAL INFORMATION

Definition and operationalization of the target variable. Standardized name of source variable.

Control variables.

Table 1 Description of the target variable RURAL/URBAN LOCALITY.

	Variable label	Variable name	Variable values
Target variable	Rural/urban locality	T_RURALURB	1 = rural 0 = urban -1 = standardized code for don't know (DK), not sure and neither/nor -9 = standardized code for non-response (NA, DK/NA) -5 = no source variable matching the target in a dataset -7 = insufficient information in description of the source variable -8 = source variable matched on the dataset level, but question not asked in a specific wave/country
Source variables	Rural/Urban locality source variable	S_RURALURB	source values (see document DETAILED TARGET VARIABLE REPORT– document's name)
Control variables	Rural/Urban respondent answer	C_RURALURB_SUBJECTIVE	1 = question answered by the respondent 0 = information comes from other sources (interviewer, sampling procedure etc.)
	Keyword 'rural' appears	C_RURALURB_KEYWORD	1=decision about coding was based on keywords 0= decision on coding was not based on keywords
	Rural/urban definition extended	C_RURALURB_UNCLEAR	1= ambiguous response category: unclear if it is rural or urban 0=corresponds to standard definition

The prefixes S, T, C are the first letters to identify the type of variables in the data set, where S stands for source variable, T for target variable and C for control variable respectively.

SURVEY PROJECTS

General description of survey projects containing source variables.

Table 2 Information on international survey projects and total number of source variables used to create target variable on RURAL/URBAN LOCALITY.

Survey Project Waves (N=5)

ARB/1, ARB/2, EB/2004, EQLS/1-3, ESS/1-5

For information on abbreviation please see <http://www.dataharmonization.org>. Waves of international survey projects correspond to the data files. In this case ARB/1 and ARB/2 are two separate data files for two waves, but ESS/1-5 is one merged data file for all five waves.

Countries (N=38)

AR BF BJ BO BR BW BZ CA CL CN CO CR CV DO DZ EC EG GH GT GY HK HN HT ID IQ JM JO JP KE KH KR LB LR LS MG ML MN MW

ISO two level country code (alpha 2). When necessary, it is extended with territory or nationality sub-codes, e.g. GB-NIR or BE-FLA.

Years (N=18)

1969, 1991-1994, 2000-2013

Exact years of survey project listed chronologically, where years 2000-2013 means that we have data from at least one survey project wave for each year within the abbreviated time span.

Source variables/questions (N=150)

See document DETAILED TARGET VARIABLE REPORT–document's name

GENERAL RULES AND PROCEDURES

a) Source data description

General description of the variety of source data. Detailed information on the source variables for each survey project (question number, questions wording, response categories, variable label and variable name) is available in DETAILED VARIABLE REPORT–RURALURB.XLSX

b) Rules of transformation of source variables into target variable

List of all harmonization rules. Example:

- if more than one aggregated variable is available, prefer the one with more detailed response category;

or

- if the size of locality is smaller than 5.000 inhabitants, code it as rural.

c) Harmonization control variables

Detailed description of control variables referring to question wording and response categories. Quality control variables for a given target variable if available.

SPECIAL CASES

Detailed information about special cases and decisions made.

COMMENTS

Additional sources used to make decisions. Information about variables excluded from analysis.

The template we present here is work in progress. We are looking forward for any comments, including about the clarity of the report's structure, its ease of understanding, and coverage of relevant information.

REFERENCES

- Data Documentation Initiative. 2014. DDI Lifecycle 3.2. <http://www.ddialliance.org/Specification/DDI-Lifecycle/3.2> [Last accessed October 13, 2015]
- Granda, Peter and Emily Blasczyk. 2010. Cross-Cultural Survey Guidelines: XIII. Data Harmonization. <http://projects.isr.umich.edu/csdi/pdf/13DataHarmonizationNov2010.pdf> [Last accessed October 13, 2015]
- Günther, Roland. 2003. "Report on Compiled Information of the Change from Input Harmonization to Ex-post Harmonization in National Samples of the European Community Household Panel – Implications on Data Quality." Wiesbaden: Statistisches Bundesamt - Working Paper 19. https://www.destatis.de/DE/Methoden/Methodenpapiere/Chintex/ResearchResults/Downloads/WorkingPaper19.pdf?__blob=publicationFile [[Last accessed October 13, 2015]
- Phillips, Drystan, Chien Sandy, Angrisani Marco, Meijer Erik and Lee Jinkook. 2014. Harmonized SHARE Documentation Version B, February 2014. <http://gateway.usc.edu/codebooks/Harmonized%20SHARE%20B.pdf> [Last accessed October 13, 2015]
- Winters, Kristi and Friedrichs Martin. 2015. QuickCharmStats 1.1 for PC Installation Guide & Software Manual. http://www.gesis.org/fileadmin/upload/dienstleistung/tools_standards/charmstats/1.1%20QCS%20Manual.pdf [Last accessed October 13, 2015]

Ilona Wysmulek is a PhD candidate at the Institute of Philosophy and Sociology, Polish Academy of Sciences (IFiS PAN) writing her dissertation on public sector corruption in Europe. She is a visiting scholar at the Mershon Center for International Security Studies, The Ohio State University and National Opinion Research Center, the University of Chicago.

Olena Oleksiyenko is a PhD candidate at the Graduate School for Social Research at IFiS PAN. Currently she works as a research assistant in the Data Harmonization Project and the Polish Panel Survey (POLPAN) 1988-2013 project conducted at IFiS PAN.

Przemek Powalko currently works in IFiS PAN for the Data Harmonization Project as a database specialist. He has presented his research on working with big data at international survey and social science conferences and workshops in Poland, United States, Germany, and Iceland.

Marcin W. Zielinski, Ph.D., works in Robert B. Zajonc Institute for Social Studies at the Warsaw University on the Polish General Social Survey and the Polish edition of the International Social Survey Programme. He is also the head of the Polish Social Data Archive (ADS).

Kazimierz M. Słomczynski is Professor Dr. Hab. at IFiS PAN and Emeritus Professor, Department of Sociology, The Ohio State University (OSU). He directs the Cross-National Studies: Interdisciplinary Research and Training program (CONSIRT) of the OSU and PAN and is Principal Investigator of the Democratic Values and Protest Behavior: Data Harmonization, Measurement Comparability, and Multi-Level Modeling project, funded by Poland's National Science Centre.