

*Harmonization:*  
Newsletter on Survey Data  
Harmonization in the Social Sciences

Volume I, Issue I  
Winter 2015

Editors  
Irina Tomescu-Dubrow  
and  
Joshua Kjerulf Dubrow  
CONSIRT

[consirt.osu.edu/newsletter/](http://consirt.osu.edu/newsletter/)

## Welcome!

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.

This newsletter is a production of [Cross-national Studies: Interdisciplinary Research and Training Program](#), of The Ohio State University (OSU) and the Polish Academy of Sciences (PAN). The catalyst for the newsletter is our ongoing 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. Słomczynski (PI), J. Craig Jenkins (PI), Irina Tomescu-Dubrow, Joshua Kjerulf Dubrow, Przemek Powalko, Marcin W. Zieliński, and research assistants: Marta Kołczyńska, Matthew Schoene, Ilona Wyszumlek, Olena Oleksiyenko, Anastas Vangeli, and Anna Franczak. For more information, please visit [dataharmonization.org](http://dataharmonization.org).

This newsletter features work from members of The Harmonization Project and the larger survey data harmonization community, including an article by Dean Lillard of The Ohio State University, and director of the Cross-national Equivalent File (CNEF). We invite all scholars interested in survey data harmonization to contribute.

## In This Issue

[History of Survey Data Harmonization, p. 2](#)

[Cross-national Equivalent File in Eight Countries, p. 4](#)

[The Harmonization Project, p. 6](#)

[Working with Harmonized Data, p. 10](#)

[Survey Data Recycling, p. 11](#)

[The Importance of Data Documentation, p. 13](#)

[Looking for Processing Errors, p. 15](#)

[Conferences and Workshops, p. 17](#)

[New Publications, p. 19](#)

[Building a Community, p. 21](#)

## Acknowledgements

Editors thank Tadeusz Krauze for comments and Marta Kołczyńska for technical assistance.

# Articles

---

## 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.

*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.*

## 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).

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

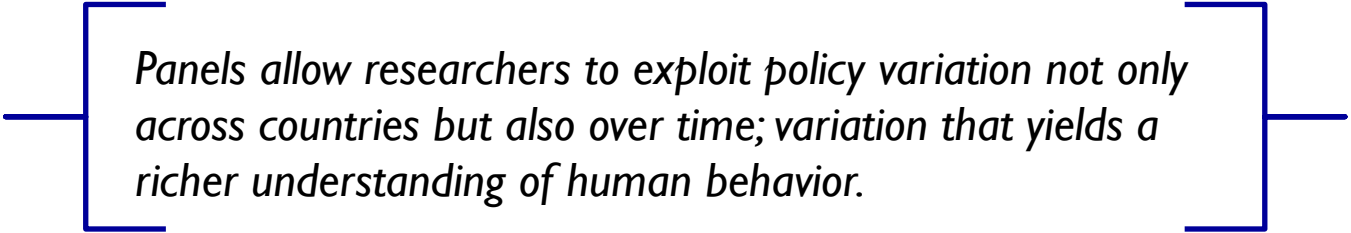
by Dean Lillard, The Ohio State University

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.<sup>1</sup> 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.<sup>2</sup>

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.



*Panels allow researchers to exploit policy variation not only across countries but also over time; variation that yields a richer understanding of human behavior.*

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](mailto:CNEF@osu.edu) or visit the home page at <http://cnef.ehe.osu.edu/>

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.

## 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.

## **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
<b>Total</b>		<b>1966-2013</b>	<b>89</b>	<b>81</b>	<b>1726</b>	<b>2293337</b>

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.<sup>1</sup> 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.<sup>2</sup> 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.

*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.

## References

- Benson, Micelle 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](http://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](http://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.
- Powalko, 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](http://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 Kolczyńska. 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](http://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.

## 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.

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

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.

*About authors, see Tomescu-Dubrow and Slomczynski in this Newsletter.*

## **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

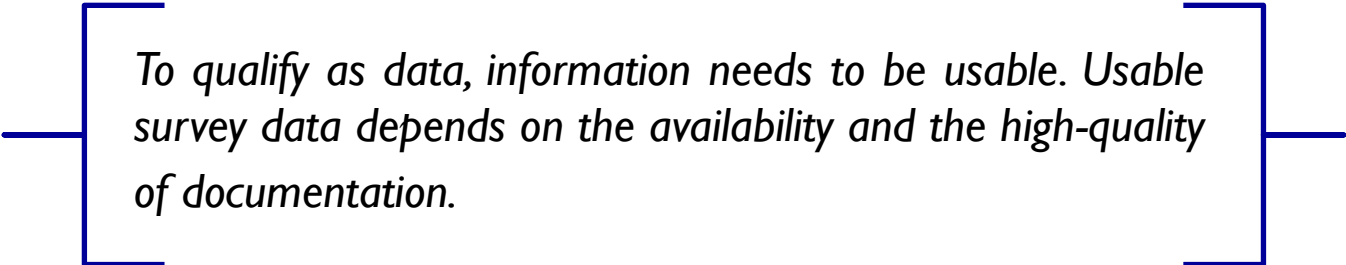
researchers, additional aspects of the survey process, such as response rates and control of fieldwork, are also important to review in order to assess survey data quality.

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 Kolczyńska 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.

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

## 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 Kolczyńska. 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.

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

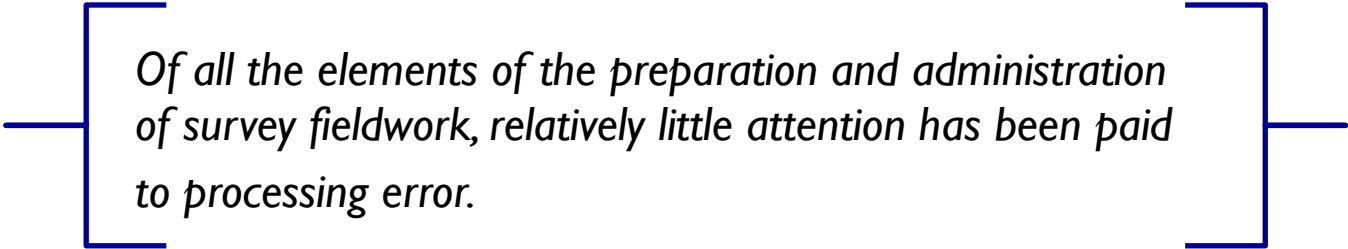
by Ilona Wysmulek, Olena Oleksiyenko, and Anastas Vangeli, Polish Academy of Sciences

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.

*Ilona Wyszumutek, 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.*



## 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.

# Conferences and Workshops

---

## Political Behavior and 'Big Data': A Mershon Center and CONSIRT Event

by Joshua Kjerulf Dubrow and Irina Tomescu-Dubrow, Polish Academy of Sciences and CONSIRT

From the wealth of quantitative cross-national survey projects, data and research, scholars know a lot about the relationship between political behavior and democracy. Yet, we have only begun to use this wealth to its fullest capacity: international survey projects could be harmonized *ex post* and turned into 'big data' that consist of an unusually large number of variables and with individuals nested in countries and time periods.

This idea informed the two-part event Interdisciplinary Studies of Political Behavior and the Use of 'Big Data', within the Harmonization Project (Polish National Science Centre grant 2012/06/M/HS6/00322; see Tomescu-Dubrow and Slomczynski in this *Newsletter*). The Mershon Center for International Security Studies at OSU and CONSIRT - Cross-national Studies: Interdisciplinary Research and Training program, OSU and the Polish Academy of Sciences (PAN), were the main organizers. They benefitted from the support of the OSU Polish Studies Initiative and of the OSU Department of Sociology.

The conference *Interdisciplinary Studies of Political Behavior: From Elections to Protests* (May 6-7, 2014 the Mershon Center) was the first part of the event. Scholars from sociology, political science, international relations, area studies and communications presented theoretically grounded empirical papers that engaged with causes and consequences of different types of political behavior. Big data methodological topics included post data harmonization of (at that point) 21 cross-national survey projects from the 1960s to the present (Kazimierz M. Slomczynski, Director of CONSIRT and Irina-Tomescu-Dubrow, associate professor at PAN, and Program Manager of CONSIRT) and political event counts from international news wires (Thomas Maher, OSU Sociology and J. Craig Jenkins, Director of Mershon). Substantive topics included a keynote speech on women's political underrepresentation worldwide by Pamela Paxton (University of Texas-Austin, Sociology), electoral protests in the developing world (Emily Beaulieu, University of Kentucky, Political Science), political legitimacy in weak states (Irfan Nooruddin, OSU Political Science), towards a unified theory of political conflict (Edward Crenshaw, OSU Sociology), democratic mobilization and Internet Communication Technologies use (Erik Nisbet, OSU School of Communication), determinants of the Arab Spring (Jenkins and colleagues), the Comparative National Elections Project (Richard Gunther and Paul Beck, OSU Political Science), Muslim women's political representation in Western Europe (Melanie Hughes, OSU alumni and assistant professor at the University of Pittsburgh, Sociology), and exit, voice and loyalty in contemporary Poland (Andrzej Rychard, Director of the Institute of Philosophy and Sociology, PAN).

The Workshop "Comparability of Survey Data on Political Behavior: Ex Post Harmonization of Selected Survey Projects" (May 8-9, 2014, Department of Sociology) focused on key methodological and statistical issues in comparability of survey data in the context of harmonization. The first day of the Workshop dealt with ways of assessing comparability of cross-national survey data for the purposes of harmonization. Established scholars, including Professor Dean Lillard, Department of Human Sciences, OSU, who directs and manages the Cross-national Equivalent File (CNEF) contributed expert insights. The second day was tailored to graduate students interested in comparative methodology of cross-national surveys.

To train the next generation of international, interdisciplinary scholars, graduate students from the Graduate School for Social Research at PAN and from departments across OSU attended both the conference and the workshop.

Event organizers have planned a special guest edited issue of the *International Journal of Sociology* based on the event "Political Behavior and Big Data." Together with this *Newsletter*, it will contribute to building a scholarly community on the methodology of survey data harmonization.

# New Publications

---

## Books

[Life-Course Smoking Behavior and National Context in Ten Countries](#) by Dean R. Lillard and Rebekka Christopoulou, Editors. Oxford University Press. Forthcoming (2015).

Co-author Dean R. Lillard describes the book:

“This edited book describes and discusses life-course smoking trajectories of seven cohorts of men and women from the United States, United Kingdom, Australia, Canada, China, Germany, Russian Federation, Spain, Turkey, and Ukraine. For almost every country the smoking trajectories are based on nationally representative samples and span more than 80 years of history - from the 1920s to the 2000s. It uses from Australia (HILDA), UK (BHPS), Germany (SOEP), and Russia (RLMS-HSE).

Several aspects of this book set it apart from existing work. It is the first book that does not rely on snapshot descriptions of smoking behavior but instead uses data on individual's smoking behavior over their whole lives. Most other descriptions rely on cross-sectional data that do not track changing smoking habits of individuals across the life-course.

The book embeds the smoking patterns in rich contexts that differ greatly across countries at a particular time and over time in a given country. It documents selected aspects of the economic, political, social, and cultural events that occurred in each country that likely affected smoking behavior of the cohorts studied. For some countries, the book presents data on real cigarette prices and taxes that have never been published.

The structure of the book facilitates comparison of smoking patterns over time and across generations, genders, and countries. In a set of country-specific chapters, the authors present smoking patterns in standardized figures and tables. In the last three chapters, the authors directly compare and contrast, across countries, the smoking behavior of men, women, and men relative to women. Finally, they provide preliminary statistical evidence about factors that might explain the smoking patterns, aiming to provoke new empirical research.”

[Online Panel Research: A Data Quality Perspective](#) by Mario Callegaro, Reginald P. Baker, Jelke Bethlehem, Anja S. Göritz, Jon A. Krosnick, and Paul J. Lavrakas, Editors. Wiley Series in Survey Methodology (2014).

From a data quality perspective, this work is relevant to survey data harmonization. The publisher describes the book:

“This edited volume is one of the first attempts to carefully examine the quality of the survey data being generated by online samples. It describes some of the best empirically-based research on what has become a very important yet controversial method of collecting data. *Online Panel Research* 19

chapters of previously unpublished work addressing a wide range of topics, including coverage bias, nonresponse, measurement error, adjustment techniques, the relationship between nonresponse and measurement error, impact of smartphone adoption on data collection, Internet rating panels, and operational issues.”

## CONSIRT Working Papers Series

CONSIRT announces a new Working Papers Series in cross-national social science research and invites scholars interested in cross-national survey data harmonization to consider submitting a working paper.

CONSIRT Working Papers are high quality, cross-national, comparative, English language scholarly work that communicates new ideas and has clear contributions to the social sciences. We see a need for scholars to access the latest in cross-national comparative research. Our mission is to promote these works in an open, electronic forum for the benefit of the international social science community.

The Working Papers Series is co-edited by [Dr. Malgorzata Mikucka](#), Université Catholique de Louvain, Laboratory for Comparative Social Research, and CONSIRT, and [Dr. Joshua Kjerulf Dubrow](#), PAN and Project and Labs Coordinator, CONSIRT.

Working Papers are available on the CONSIRT website, [consirt.osu.edu/working-papers-series](http://consirt.osu.edu/working-papers-series). Posting a Working Paper on [consirt.osu.edu](http://consirt.osu.edu) does not preclude simultaneous or subsequent publication elsewhere, including other working papers series.

### Current Working Papers include:

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 Papers Series 1 (CONSIRT Labs: Methodology of Survey Data Harmonization)* at [consirt.osu.edu](http://consirt.osu.edu).

Powalko, Przemek. 2014. “Working with Big Data: Experiences with the Cross-national Survey Data Harmonization Project.” *CONSIRT Working Papers Series 2 (CONSIRT Labs: Methodology of Survey Data Harmonization)* at [consirt.osu.edu](http://consirt.osu.edu).

Schoene, Matthew and Marta Kolczyńska. 2014. “Survey Data Harmonization and the Quality of Data Documentation in Cross-National Surveys.” *CONSIRT Working Papers Series 3 (CONSIRT Labs: Methodology of Survey Data Harmonization)* at [consirt.osu.edu](http://consirt.osu.edu).

## Harmonization would like to hear from you!

We created this *Newsletter* to share news and help build a growing community of those who are interested in harmonizing social survey data. We invite you to contribute to this Newsletter. Here's how:

### 1. Send us content!

- Send us your announcements (100 words max.), conference and workshop summaries (500 words max.), and new publications (250 words max.) that center on survey data harmonization in the social sciences;
- Send us your short research notes and articles (500 – 1000 words) on survey data harmonization in the social sciences. We are especially interested in advancing the methodology of survey data harmonization. If we have any questions or comments about your items, we will work with you to shape them for this *Newsletter*.

Send it to: Joshua Kjerulf Dubrow, [dubrow.2@osu.edu](mailto:dubrow.2@osu.edu).

### 2. Tell your colleagues!

To help build a community, this *Newsletter* is open access. We encourage you to share it in an email, blog or social media (Facebook, Twitter, Google+, and so on).

Here's text for an email that you can send to your colleagues!

*Dear Colleagues,*

*There's a new publication that might interest you. It's called, *Harmonization: Newsletter on Survey Data Harmonization in the Social Sciences*, from *The Ohio State University and the Polish Academy of Sciences*.*

*Harmonization of survey data has many uses, and has great potential to advance methodology in the social sciences. The Harmonization team wants to help build a community of scholars, institutions and government agencies who work on harmonizing social survey data.*

*The first issue is available at [consirt.osu.edu/newsletter/](http://consirt.osu.edu/newsletter/)*

*Enjoy!*

## Support

This issue was partly supported by a grant from the (Polish) National Science Centre (2012/06/M/HS6/00322).

## Copyright Information

*Harmonization: Newsletter on Survey Data Harmonization in the Social Sciences* is copyrighted under Creative Commons Attribution-NonCommercial-ShareAlike 3.0 United States ([CC BY-NC-SA 3.0 US](https://creativecommons.org/licenses/by-nc-sa/3.0/us/)).

“You are free to: Share — copy and redistribute the material in any medium or format; Adapt — remix, transform, and build upon the material. The licensor cannot revoke these freedoms as long as you follow the license terms. Under the following terms: Attribution — You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use. NonCommercial — You may not use the material for commercial purposes. ShareAlike — If you remix, transform, or build upon the material, you must distribute your contributions under the same license as the original. No additional restrictions — You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.”