

Harmonization:
Newsletter on Survey Data
Harmonization in the Social Sciences

Volume 1, Issue 2

Fall 2015

Editors

Irina Tomescu-Dubrow

and

Joshua Kjerulf Dubrow

CONSIRT

consirt.osu.edu/newsletter/

ISSN 2392-0858

Building a Community

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. Wyszynski 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.

In This Issue

[Harmonized Survey Datasets and Their Sources, p. 2](#)

[Harmonization of Educational Attainment Variables, p. 5](#)

[The Barometer Surveys: Quality of Political Trust Items, p. 7](#)

[The CharmStats Program for Survey Data Harmonization, p. 12](#)

[Standardization: Target Variable Report Template, p. 13](#)

[Conferences and Workshops, p. 18](#)

[News, p. 19](#)

[Contact Us, p. 20](#)

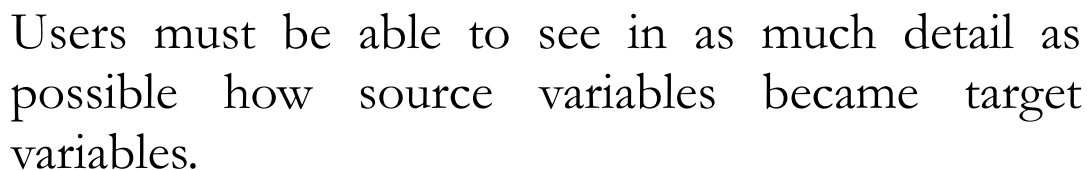
Acknowledgements
Editors thank Marta Kołczyńska for technical assistance.

Articles

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.



Users must be able to see in as much detail as possible how source variables became target variables.

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.

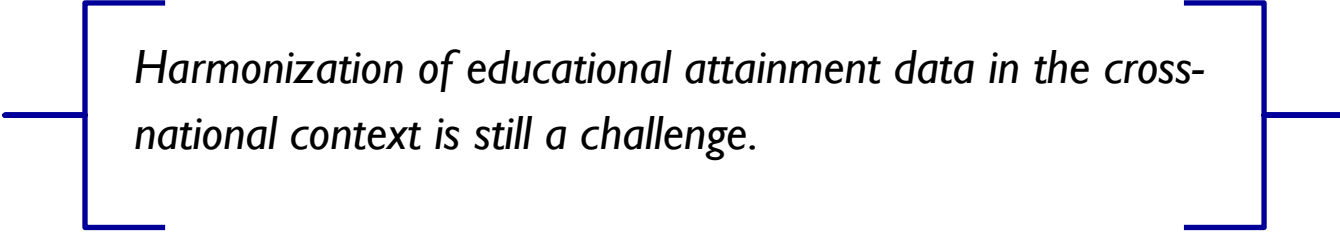
Peter Granda is Associate Director of the Inter-university Consortium for Political and Social Research (ICPSR). He also directs both the General Archive and the Health and Medical Care Archive, a topical archive supported by the Robert Wood Johnson Foundation. He has a long association with the cultures 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).



Harmonization of educational attainment data in the cross-national context is still a challenge.

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 and 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, Verena and Silke L. Schneider. 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 in spite of 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).

Political trust items are generally administered as a battery of items, but the sequence differs by international survey project.

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

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).

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 (see Table 1).

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; see Table 1). 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

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.

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.

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].

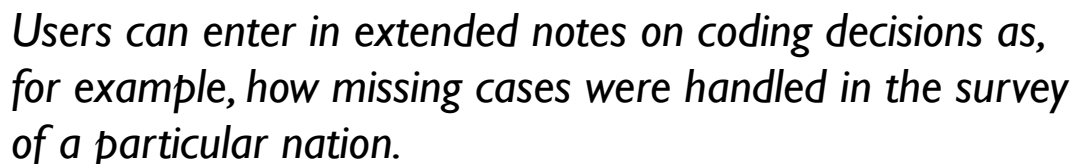
Wiebke Breustedt is Ph.D. student at the University of Cologne (Germany). She worked as research assistant at the Chair for Comparative Politics at the University of Duisburg-Essen (Germany) and as lecturer at the Chair for Comparative Politics at the University of Cologne. Her research and publications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

Global Barometer Survey (Module 1)	Afrobarometer (2005)		Asian Barometer (2005-2008)		Latinobarómetro (2003)		Arab Barometer (2006-2007)	
Number and labels of response categories	Question wording	Number and labels of response categories	Question wording	Number and labels of response categories	Question wording	Number and labels of response categories	Question wording	Number and labels of response categories
1 (none at all) 2 (not very much trust) 3 (quite a lot of trust) 4 (a great deal of trust) 7 (do not understand the question) 8 (can't choose) 9 (decline to answer) -1 (missing) 0 (not applicable)	How much do you trust each of the following, or haven't you heard enough about them to say	0 (not at all) 1 (just a little) 2 (somewhat) 3 (a lot) 9 (don't know/haven't heard enough) 98 (refused to answer) -1 (missing data)	I'm going to name a number of institutions. For each one, please tell me how much trust do you have in them? Is it a great deal of trust, quite a lot of trust, not very much trust, or none at all?	1 (a great deal of trust) 2 (quite a lot of trust) 3 (not very much trust) 4 (none at all) 7 (do not understand the question) 8 (can't choose) 9 (decline to answer)	Please look at this card and tell me how much confidence you have in each of the following groups/institutions. Would you say you have a lot, some, a little or no confidence?	1 (have a lot) 2 (some) 3 (a little) 4 (no confidence) 8 (don't know) 0 (no answer)	I'm going to name a number of institutions. For each one, please tell me how much trust you have in them. Is it a great deal of trust, quite a lot of trust, not very much trust, or none at all?	1 (a great deal of trust) 2 (quite a lot of trust) 3 (not very much trust) 4 (none at all) 98 (can't choose/don't know) 99 (decline to answer)
Objects of trust	Afrobarometer (2005)		Asian Barometer (2005-2008)		Latinobarómetro (2003)		Arab Barometer (2006-2007)	
	Item wording							
Prime Minister or President	The President/ Prime Minister (q55a)		The Presidency or Prime Minister (qll07)		The President (p23ste)		Prime minister (q2011)	
The Courts	Courts of Law (q55i)		The Courts (q007)		The Judiciary (p21ste)		The Courts (q2012)	
The National Government	n.a.*		The national government [in capital city] (q008)		The Government (p23stg)		n.a.	
Political Parties	n.a.*		Political parties [not any specific party] (q009)		The Political Parties (p21std)		Political Parties (q2015)	
The Parliament	The Parliament/ National Assembly (q55b)		Parliament (q010)		The National Congress/Parliament (p21stf)		Parliament (q2013)	
The Military	The Army (q55g)		The military (or armed forces) (q012)		The Armed Forces (p21stg)		n.a.	
The Police	The Police (q55h)		The police (q013)		The Police (p21stb)		The Police (q2014)	
The Local Government	Your Elected Local Government Council (q55d)		Local government (q014)		n.a.		n.a.	
The Election Commission	The Electoral Commission (q55c)		The election commission [specify institution by name] (q017)		n.a.		n.a.	
<p><i>Notes:</i> the variable labels are indicated as they appear in the merged data files; the question wording, number and labels of response categories are indicated as they appear in the English language master questionnaires and the GBS data file; * the Afrobarometer (2005) includes items measuring 'trust in the ruling party' and 'trust in the opposition political parties'</p> <p><i>Sources:</i> own compilation based on the survey data provided online at afrobarometer.org; asianbarometer.org; arabbarometer.org; latinobarometro.org; globalbarometer.net</p>								

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 Wysmułek, Olena Oleksiyenko, Przemek Powalko, Marcin W. Zieliński,
Kazimierz M. Słomczyński

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.

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 variables: $T = f(S)$.

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

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.

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 subcodes, 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 Wyszumłek 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 Oleksijenko 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. Zieliński, 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. Slomczyński 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.

Conferences and Workshops

The “Cross-national Survey Harmonization and Analysis: Weights, Data Quality and Multi-level Modeling” Workshop at The Ohio State University

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

In 2015, Cross-national Studies: Interdisciplinary Research and Training program (CONSIRT.osu.edu) at The Ohio State University (OSU) and the Polish Academy of Sciences organized the international workshop “Cross-national Survey Harmonization and Analysis: Weights, Data Quality and Multi-level Modeling.” The workshop was held over six days (May 11 - 14, and May 15 – 16) at the OSU Mershon Center for International Security Studies and the OSU Department of Sociology.

Following one of CONSIRT’s main objectives – to enhance graduate students’ educational and professional training and facilitate their international collaboration – the workshop taught students how to employ survey weights, information on the quality of survey data, and multilevel modeling to address methodological and substantive problems in quantitative social science research. In attendance was an interdisciplinary mix of twelve graduate students, including ten from OSU Sociology, Political Science, and Communication, and two from the Graduate School for Social Research of the Polish Academy of Sciences. Participation was free of charge.

The workshop featured a dataset of European countries and the U.S., constructed via ex-post harmonization of individual-level measures of political behavior and demographics selected from various international survey projects, including the World Values Survey, International Social Survey Program, and the European Social Survey, and appended with country-level variables from non-survey sources and a range of quality-control indicators. This dataset is part of the large-scale database of 2.3 million respondents, covering a total of 142 countries and territories, and spanning almost the period 1966-2014, built within the Survey Data Recycling project (detailed in Harmonization Newsletter v1 no1 2015).

The opening lecture on data harmonization and an overview of the Survey Data Recycling framework was delivered by Professor Kazimierz M. Slomczynski, of OSU and the Polish Academy of Sciences, and Director of CONSIRT, and Dr. Irina Tomescu-Dubrow, Associate Professor at the Polish Academy of Sciences and Program Coordinator of CONSIRT.

The workshop had three parts. Part one was “Survey Weights in Comparative Analysis” (May 11-12). Sampling weights are needed to make valid inferences from samples to the populations from which they were drawn. In cross-national research oftentimes one needs to deal with two types of population: national population, and the population of a set of countries that we want to include in common analyses. This workshop reviewed the major types of survey weights used in face-to-face surveys and went in-depth on weights for cross-national surveys. The focus was on post-stratification weights and population weights. Presenters included Marcin Zielinski of the Polish Social Data Archive and the University of Warsaw, and Przemek Powalko of the Polish Academy of Sciences.

The second part was “Data Quality in Cross-National Analysis” (May 13-14). International projects differ in terms of quality of their data documentation, data processing, and the actual data contained in the computer file. Workshop participants learned how to account for such variation in data quality using quality-control variables constructed for each of these three aspects. They assessed empirically how these control indicators impact the accuracy of statistical results involving substantive variables. Presenters included Ilona Wysmulek and Olena Oleksiyenko of GSSR, and Marta Kolczyńska and Matt Schoene of OSU.

The third part was “Multi-Level Modeling” (May 15-16). Cross-national analyses oftentimes call for individual-level variables as well as contextual measures such as country characteristics. Hierarchical multi-level models are a good solution for modeling well-defined nested levels. At the same time, data quality controls are, or could be, defined on the level of national surveys or even entire international survey projects, which calls for complex multi-level modeling techniques. This workshop discussed both hierarchical and non-hierarchical multi-level modeling. The primary lecturer was Professor Robert Kunovich of the University of Texas- Austin and OSU alumni.

The event was co-organized by CONSIRT, the OSU Mershon Center for International Security Studies, and with the financial support of the OSU Departments of Political Science and Sociology, and the Polish Studies Initiative at OSU. It benefited from grants from the Polish National Science Centre (NCN): the international cooperation *Harmonia* grant 2012/06/M/HS6/00322 (Principal Investigator: Kazimierz M. Slomczynski) and *Preludium* grant 2012/05/N/HS6/03886 (Principal Investigator: Marta Kolczyńska).

Joshua Kjerulf Dubrow and Irina Tomescu-Dubrow are Associate Professors at IFiS PAN and are administrators of CONSIRT.

New Publications

Ask: Research and Methods Re-publishes the Harmonization Newsletters of 2015

Ask: Research and Methods (ASK) is an open access, peer-reviewed academic journal devoted to the methodology of social science research. Since 1995, its mission has been to publish innovative work in the social sciences that focuses on methodological issues inherent to all phases of social research. Since 2014, it is co-published by The Ohio State University and the Institute of Philosophy and Sociology of the Polish Academy of Sciences. This co-publishing agreement was made possible by the organizational support of Cross-national Studies: Interdisciplinary Research and Training Program.

CONSIRT is pleased to announce that the first two issues of Harmonization will be re-published in ASK. All volumes of ASK, 1995 to the present, including each individual article, are archived in The Ohio State University’s Knowledge Bank. The Harmonization Newsletter will be re-published as a whole.

You can find more about ASK at askresearchandmethods.org and CONSIRT.osu.edu/askresearchandmethods.

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 provide feedback, and to contribute to this Newsletter. Here's how:

1. Send us content!

- 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 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 it to: Joshua Kjerulf Dubrow, 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 current issues are available at consirt.osu.edu/newsletter/

Enjoy!

Support

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. Slomczynski (PI), J. Craig Jenkins (PI), Irina Tomescu-Dubrow, Joshua Kjerulf Dubrow, Przemek Powalko, Marcin W. Zieliński, and research assistants: Marta Kolczyńska, Matthew Schoene, Ilona Wysmulek, Olena Oleksiyenko, Anastas Vangeli, and Anna Franczak. For more information, please visit dataharmonization.org.

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](http://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.”