gesis Leibniz Institute for the Social Sciences

Meet the Experts

"Ex Post-Harmonised" Comparative Survey Data Files at GESIS - and What to Expect Of Them

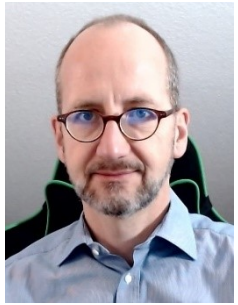
Markus Quandt & Ivet Solanes-Ros

GESIS Meet-the-Experts Series

June 01 2023

Leibniz Association

Speakers/Authors



Dr. Markus Quandt

- Head of Team International Studies in Department Survey Data Curation at GESIS
- PhD in Economics, Social Science focus
- Comparative survey analysis, equivalence, research data management
- Contact: markus.quandt@gesis.org

Contributing: Ivet Solanes Ros

- Data Processing Specialist in SDC Teams „International Studies“ and „Survey Data Augmentation“
- Dipl.-Soz.
- Survey data processing with European Values Study and many other comparative surveys
- Contact: ivet.solanes@gesis.org

gesis Leibniz Institute for the Social Sciences

Meet the Experts

Logistics

- This talk will be recorded.
  We do not record the Q&A session after the talk.
- Participants are muted during the session. Questions will be collected during presentation and answered after the talk.
- Please use the private chat function and send questions to the account "Q&A host (MTE)".
  If you post in the general chat, your name and message will be visible to all participants. Of course, this is also possible; we kindly ask you to prefer the private chat to "Q&A host (MTE)" while the presentation is going on.
- Recording and slides will be made publicly available on the GESIS website and on our YouTube channel.

gesis Leibniz Institute for the Social Sciences

Meet the Experts

"Ex Post-Harmonised" Comparative Survey Data Files at GESIS - and What to Expect Of Them

Markus Quandt & Ivet Solanes-Ros

GESIS Meet-the-Experts Series

June 01 2023

Leibniz Association

gesis Leibniz Institute for the Social Sciences

Scope of This Talk

- Practical aspects of using cross-national and trend data at GESIS
  - I. (2nd-tier) harmonised cross-cultural data: why and what for?
  - II. How to find and get harmonised data at GESIS?
  - III. An overview of what to find at GESIS
  - IV. Challenges in Data Harmonisation
  - V. Example(s)

gesis Leibniz Institute for the Social Sciences

I. Tiers of Harmonisation

- **Ex-ante harmonisation** – in advance design of measurement processes for different data sources such that maximum equivalence can be expected.
- **Ex-post harmonisation** – harmonise across data sources for which ex-ante harmonisation was not (fully) in place.
  In terms of *representational measurement theory*: try to remove any extant differences among ‚segments' of the data (=„*numerical relational structure*") that are meant to be describing the same „*empirical relational structure*".
- Different „sources" constitute such ‚segments' ≈ different populations, different samples and/or different instruments, or *any other outcome-relevant differences in the data collection/measurement process.*
- **2nd-tier harmonisation** occurs when we harmonise across *different types of sources in sequence*, i.e., when we combine datasets that have already been internally harmonised for a certain type of source differences, but that now require additional harmonisation steps due to the new level of combination.

gesis Leibniz Institute for the Social Sciences

I. Applications of Ex-Post Harmonised Data

1st tier:

- Across nations/regions/populations –> **Cross-national comparison** (with both ex-ante and ex-post harmonisation approaches)
- Within-nation (more generically – within the same population), same time, across-samples -> **increase sample sizes by data pooling**, e.g. when observing small groups
- Across time (with different samples) -> **Trend analyses**, time-series cross-sectional research designs

2nd tier:

- Across time & nations -> **Comparative trend analyses**, look at trajectories for different world regions/cultural areas…
- Across (comparative) survey programs -> e.g., pooling *International Social Survey Programme* data with *World Values Survey* data to **increase regional or temporal coverage**
- (Across macro-data sources, or for **linking** survey data with macro data -> identifiers such as regions or category codes may need ‚harmonisation')

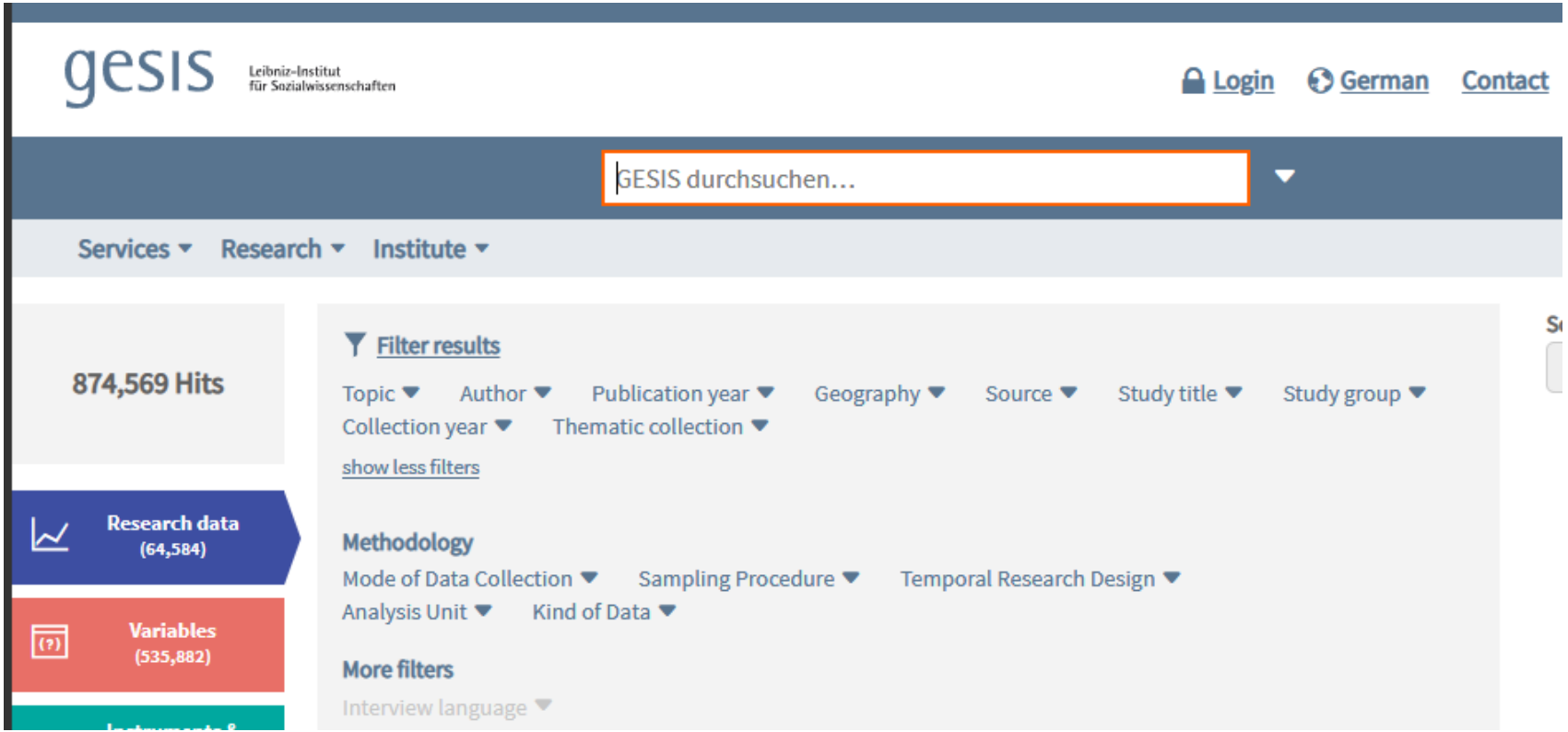
gesis Leibniz Institute for the Social Sciences

II. How to Find Harmonized Data at GESIS

- No systematic search options yet
  - Use generic search at https://search.gesis.org/
  - Curated collections often have harmonisation support
  - Some designated harmonisation support projects

- …but some guidance is coming. Watch this space: https://www.gesis.org/en/services/processing-and-analyzing-data/data-harmonization

gesis Leibniz Institute for the Social Sciences



gesis Leibniz Institute for the Social Sciences

III. Examples of What to Find at GESIS

- National Surveys Harmonized by GESIS staff
  - ALLBUS Cumulations (1980 – 2018) – data from all ALLBUS cross-section samples , overview of versions https://www.gesis.org/en/allbus/download/download-kumulationen
  - Politbarometer Cumulations – Annual data from all monthly Politbarometer cross-section samples; long-term trend file 1977-2021, https://search.gesis.org/research_data/ZA2391?doi=10.4232/1.13837
  - GLES Querschnitt 2009-2017, Kumulation (only in German) https://www.gesis.org/gles/dokumentation/einzelne-datensaetze/gles-querschnitt-2009-2017-kumulation
  - 2nd tier: Question Link: Script collection to apply elaborately calibrated harmonisations to instruments from selected German survey samples, https://www.gesis.org/en/services/processing-and-analyzing-data/data-harmonization/question-link
  - MISSY to help the handling of official statistics data, here for metadata and scripts to work with the German Mikro-Zensus: https://www.gesis.org/missy/ , in particular GESIS Mikrozensus Trend file 1962-2016, https://www.gesis.org/gml/mikrozensus/mikrozensus-trendfile
- International Surveys Harmonized by GESIS staff
  - ISSP (Internations Social Survey Programme) Cumulations – we strive to produce cumulations for each topical module with at least three reiterations, see overview pages here: https://www.gesis.org/en/issp/modules/data-cumulations
  - 2nd tier: CSES IMD – Comparative Study of Electoral Systems Integrated Module Dataset, https://cses.org/data-download/cses-integrated-module-dataset-imd/
  - EVS – European Values Study Trendfiles, https://europeanvaluesstudy.eu/methodology-data-documentation/evs-trend-file-1981-2017/ and https://search.gesis.org/research_data/ZA7503?doi=10.4232/1.14021
  - 2nd tier: European Values Study & World Values Survey „Integrated Values Surveys (IVS) 1981-2022“ - https://europeanvaluesstudy.eu/methodology-data-documentation/integrated-values-surveys/
  - MISSY again, here for metadata and other help with EU-Labour Force Survey and EU-SILC (Income and Living Conditions): https://www.gesis.org/missy/

gesis Leibniz Institute for the Social Sciences

III. Examples of What to Find at GESIS

- Single-shot project-based (often with external partners):
  - „Landtagswahlen – Integrierter Datensatz 1962 - 2004“ – surveys on federal state elections in Germany, https://search.gesis.org/research_data/ZA4182?doi=10.4232/1.12389
  - 2nd tier: CARPE – Church Attendance and Religious Change in Europe, AGGREGATE LEVEL TSCS data derived from several comparative surveys, https://search.gesis.org/research_data/SDN-10.7802-2040
  - 2nd tier: HaSpaD – German Panel Studies on Partnership Relations, harmonisation scripts, https://www.gesis.org/angebot/daten-aufbereiten-und-analysieren/datenharmonisierung/harmonisierung-von-paarbiografien and https://doi.org/10.7802/2429
  - 2nd tier: ONBound – Survey and Context Data on National Identities and Religion, harmonisation scripts, https://www.gesis.org/en/services/processing-and-analyzing-data/data-harmonization/onbound
  - 2nd tier: COORDINATE – Child well-being data harmonised across European national (panel) surveys, scripts planned, publication pending
- In the general data archive, compiled by (mostly) external researchers:
  - We don‘t always know…
  - 2nd tier: Mannheim Eurobarometer Trend File 1970-2002 (Schmitt/Scholz/Moschner), latest version here: https://search.gesis.org/research_data/ZA3521
  - 2nd tier: Harmonized Eurobarometer 2004-2021 (Russo/Bräutigam), https://search.gesis.org/research_data/SDN-10.7802-2539

gesis Leibniz Institute for the Social Sciences

IV. Some Commonalities of GESIS-produced materials

- Ready-to-use datasets available for many intensily curated survey programs
- 2nd-tier projects prefer provision of harmonisation scripts instead of datasets
- Focus on usability and complete documentation/transparency
- Careful mapping for categorical variables and linkage keys, coding towards (international) classifications (ISO country codes or NUTS, employing political party classifications…)
- Harmonisation methodology for quantitative measures limited to normalisation and standardisation, e.g. via ‚linear stretching' (exception: QuestionLink)

gesis Leibniz Institute for the Social Sciences

IV. Common Challenges in Data Harmonisation

- Instrument/variable level:
  - **Normalizing and standardising scales to common ranges and units**
  - **Establish equivalence across measures (reliability and validity)**
  - Instrument application errors
  - Coding errors (in the original data, AND in the harmonised data!)
- Study/survey level:
  - Variation in definitions of population
  - Variation in sampling frames/designs
  - Non-obvious variation in data collection timing (with diff. samples for same population); deviation of fieldwork from nominal wave years
  - ‚House effects' – hidden quality differences in data collection processes
  - Inconsistent availability of weight variables
- Documentation gaps
- Most challenges above do not fit well with standard conceptions of data quality used by measurement theory – they concern qualitative aspects and face validity.
- Many ad hoc harmonisation projects only look at the standardization of scales!

V. Two Examples

- a. European Values Study Trend file and World Values Survey Trend file, integrated to the „Integrated Values Surveys“
- b. A research article, using data from ONBound

gesis Leibniz Institute for the Social Sciences

V. EVS and WVS Joint Trend File 1981-2017

- Cumulates over 5 EVS and 7 WVS campaigns;
  - *Initially* based on their cross-sectional comparative datasets;
  - *Then* separately converted to EVS- and WVS- trend files;
  - *Then* those trend files to be merged via a script executed by the user.

- Result:
  - Time span 1981 to 2022
  - 122 countries
  - 424 country*year cells (occas. covered both by EVS and WVS!)
  - 838 variables (only variables that were repeated at least once are included)
  - 663,965 interviews

gesis Leibniz Institute for the Social Sciences

V. EVS and WVS – Some Problems

EVS:

SHOW CARD 37 – READ OUT AND CODE ONE ANSWER PER LINE

Q37 Here are two changes in our way of life that might take place in the near future. Please tell me for each one, if it were to happen whether you think it would be a good thing, a bad thing, or you do not mind.

		Good	Bad	Don't mind	DK	NA
v113	Less importance placed on work in our lives	1	2	3	8	9
v114	Greater respect for authority	1	2	3	8	9

WVS:

I'm going to read out a list of various changes in our way of life that might take place in the near future. Please tell me for each one, if it were to happen, whether you think it would be a good thing, a bad thing, or don't you mind? (*Code one answer for each*):

		Good	Don't mind	Bad
Q43	Less importance placed on work in our lives	1	2	3
Q44	More emphasis on the development of technology	1	2	3
Q45	Greater respect for authority	1	2	3

-> Harmonisation decision: recode EVS variable to WVS pattern

gesis Leibniz Institute for the Social Sciences

V. EVS and WVS – Some Problems

C001	Jobs scarce: Men should have more right to a job than women (3 categories)	Common format for both EVS and WVS in **older** waves	In recent waves, some WVS countries still use 3-point scale
C001_01	Jobs scarce: Men should have more right to a job than women (5-point scale)		In recent waves, **most** WVS and **all** EVS members used this
C002	Jobs scarce: Employers should give priority to (nation) people than immigrants (3 categories)	Common format for both EVS and WVS in **older** waves	In recent waves, some WVS countries still use 3-point scale
C002_01	Jobs scarce: Employers should give priority to (nation) people than immigrants (5-point scale)		In recent waves, **most** WVS and **all** EVS members used this

-> Harmonisation decision: Create a 3-point scale version of variable for all instances, but also leave original 5-point scales in joint data set.

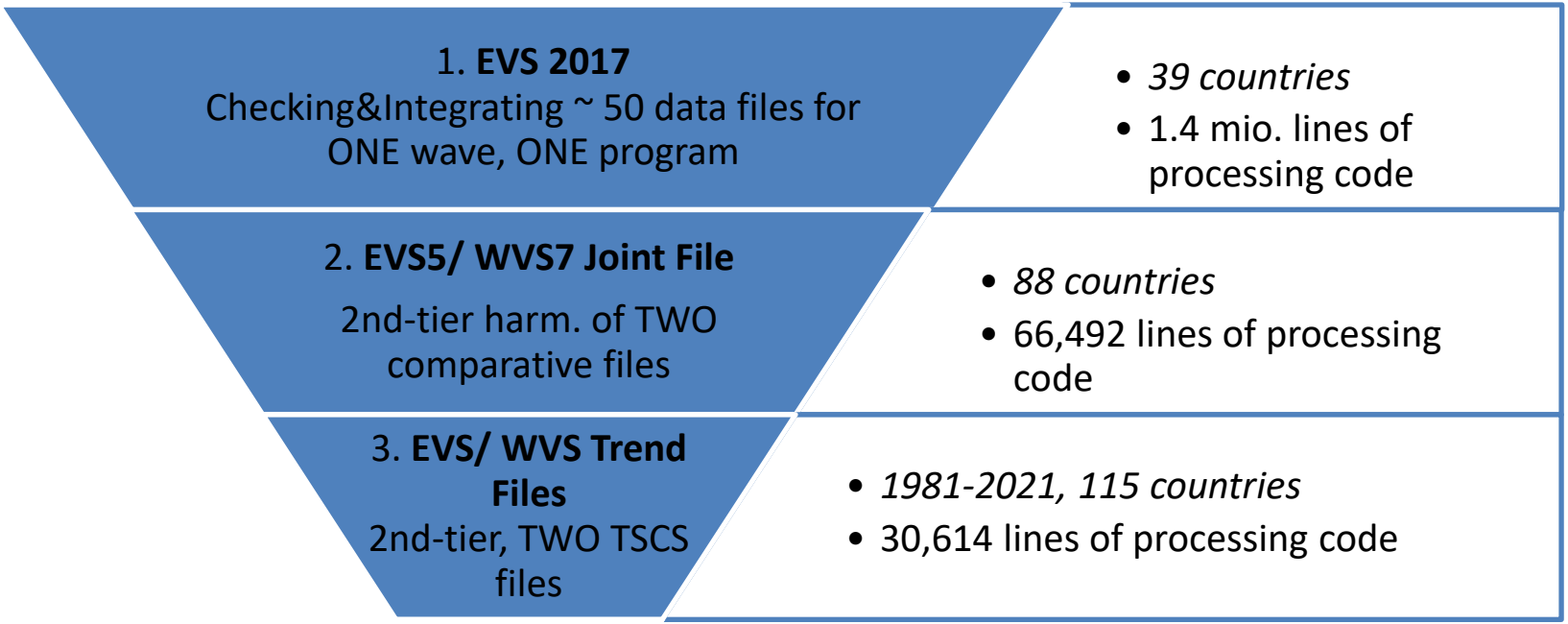
gesis Leibniz Institute for the Social Sciences

V. EVS and WVS – Some Problems

Country/Territory	EVS 2017 or WVS7	Sample size	Fieldwork period	Mode of data collection	Languages fielded	Main events between the FW Periods
Armenia	EVS	1500	20-02-2018 - 30-04-2018	PAPI	Armenian	Second Nagorno-Karabakh War, COVID
	WVS	1223	07-05-2021 - 07-06-2021	CAPI		
Czechia	EVS	1811	17-09-2017 - 03-12-2017	CAPI / PAPI	Czech	COVID, Invasion in Ukraine
	WVS	1200	11-02-2022 - 13-05-2022	CAPI		
Germany	EVS	2170	23-10-2017 - 04-04-2018	CAPI	German	
			20-09-2018 - 28-11-2018	CAWI		
			16-11-2017 - 20-03-2018	Mail		
	WVS	1528	25-10-2017 - 31-03-2018	CAPI		
Great Britain	EVS	1788	12-02-2018 - 16-07-2018	CAPI	English	Brexit, COVID, Invasion in Ukraine
	WVS	2609	02-03-2022 - 07-09-2022	CAPI / CAWI/ Mail / Video interviewing		
Netherlands	EVS	2404	31-08-2017 - 28-02-2018	CAPI	Dutch	COVID
			11-09-2017 - 30-01-2018	CAWI		
	WVS	2145	03-01-2022 - 25-01-2022	CAWI		
Romania	EVS	1613	03-02-2018 - 05-05-2018	CAPI	Romanian	
	WVS	1257	30-11-2017 - 02- 04- 2018	CAPI		
Russia	EVS	1825	07-11-2017 - 25-12-2017	CAPI / PAPI	Russian	
	WVS	1810	07-11-2017 - 29-12-2017	CAPI / PAPI		
Serbia	EVS	1499	10-11-2018 - 21-12-2018	PAPI	Serbian	
	WVS	1046	20-05-2017 - 07-07-2017	PAPI		
Slovakia	EVS	1432	26-09-2017 - 01-12-2017	CAPI	Slovak, Hungarian	COVID
	WVS	1200	19-01-2022 - 22-02-2022	CAPI	Slovak	
Ukraine	EVS	1612	02-11-2020 - 23-11-2020	PAPI	Ukrainian, Russian	
	WVS	1289	25-07-2020 - 14-08-2020	CAPI		

gesis Leibniz Institute for the Social Sciences

V. Reducing Heterogeneity?



Notes:
1. Data relate to previous release versions of all input files.
2. Code counts also includes generic curation tasks.

gesis Leibniz Institute for the Social Sciences

EVS-WVS_timeseriesplot.R | df_agg | df1

Source on Save

```
1  setwd("C:/Users/mquan/gesisbox/Projekte/MtE2023_06_Ha
2  library(tidyverse)
3  library(haven)
4  library(dplyr)
5
6  df1 <- read_sav("IVS_1981_2022_v2n.sav")
7
8  df_agg <- df1 %>%
9    group_by(S009,S020,S001) %>%
10   summarise(r = cor(A008,A009,method="pearson",use="p
11
12 plot(df_agg$meanhappy,df_agg$meanhealth,col=df_agg$S0
13
14 legend("bottomright",legend = c("EVS", "WVS"), col =
15
16
```

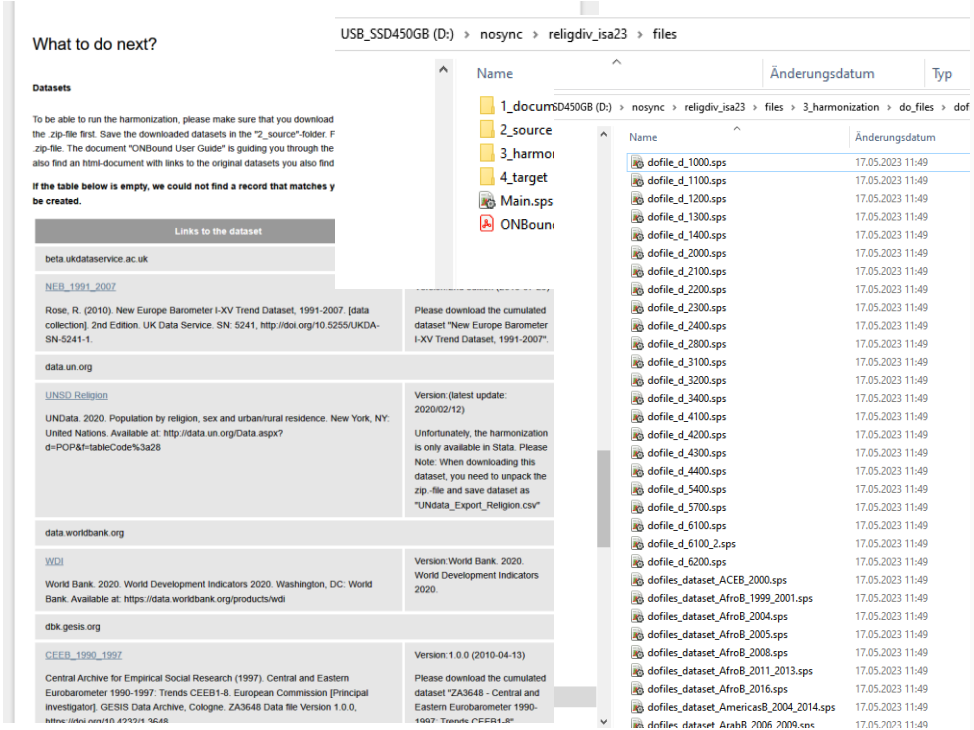
gesis Leibniz Institute for the Social Sciences

V. National Pride and Non-Religiosity

- Bechert, Insa (2012): **Of Pride and Prejudice—A Cross-National Exploration of Atheists' National Pride.** *Religions* , Vol. 12, No. 8 p. 648
- "[The] ONBound dataset (Bechert et al. 2020) is based upon data from ***seventeen survey programs*** **and** ***two macro sources*** that contain the relevant variables […]. It covers 99 countries, of which for ***35 countries***, all crucial micro and macro-level variables necessary for the Multilevel Analyses are available and a considerable number of atheists could be observed. For the time span analyzed, from ***1981 to 2018***, this selection yields almost *three million respondents*, of whom almost ***700,000 from 11 surveys*** answered a question on national pride."

gesis Leibniz Institute for the Social Sciences

V. Preparatic



File Edit Run Windows Help

```
* Encoding: UTF-8.
Begin program.
import spss, spssaux
import os, shutil

            # ------------------------------------------------------------ #
            # ONBOUND-Harmonization: Main file          #
            # ------------------------------------------------------------ #

*************************************************************.
* INFORMATION ON DOFILE
* Title: ONBound-Harmonization Datafile
* Version: v.1.0.0
* Last changes: 2020-06-30
* Author(s): May, Antonia; Werhan, Katharina; Quandt, Markus; Bechert, Insa (2020): ONBound-project team
*************************************************************.

*************************************************************.
* Content:
* This do-file sets necessary presets, defines ados, loads, changes
* and harmonizes all micro- and macro-level data sets as selected
* by the user.
*
* Downloads:
* Before running this harmonization-file, please download all
* datasets you chose. You need to following the links as indicated
* on the website (http://onbound.gesis.org/wizard) and in the ONbound_List of source datasets.xlsx
* within the downloaded folder ".../1_documentation".
* Please save the data with the original file-names and formats
* (including .zip) here: ".../ONBound/2_source"
*
* Attention: Some original datasets require additional steps:
*
* Some sources need some preparation by users
** Asian Barometer comes in .rar-format.
*   Please unpack the .rar-archive (using Win-RAR, 7-Zip or other program)
*   and save the .sav-files in the 2_source folder.
*
** UNSD2 (Religious Denomination) had to be renamed in the 2_source folder.
*   Please unzip and rename this dataset to "UNdata_Export_Religion.csv"
*
* Please read the ONBound Users Guide carefully.
*************************************************************.

*************************************************************.
*  GENERAL SECTION - MUST INCLUDE 0 - TO DO USERS = SET PATHS.
*************************************************************.

spss.Submit("""SET UNICODE ON.""")

#Set directory for harmonization process. All further commands will be based on this directory.
#Has to be set for SPSS and Python to the same folder.

#SPSS.
spss.Submit("""cd "I:\ONBound\SPSS".""")     #CHANGE THIS PATH TO MATCH THE LOCATION OF THE ONBOUND-FOLDER ON YOUR MACHINE!.
# Python.
wd = "I:\ONBound\SPSS/"                      #CHANGE THIS PATH TO MATCH THE LOCATION OF THE ONBOUND-FOLDER ON YOUR MACHINE!.

# ATTENTION: The path in line 55 should not end with "/" wheras the path in line 57 has to end with "/".

*************************************************************.
* FILTER SECTION - SETTING FILTERS AS DEFINED BY USER THROUGH WIZARD.
*************************************************************.

datasets_macro = ["CARPE","WRP","RDI","RCS"]
```

gesis Leibniz Institute for the Social Sciences

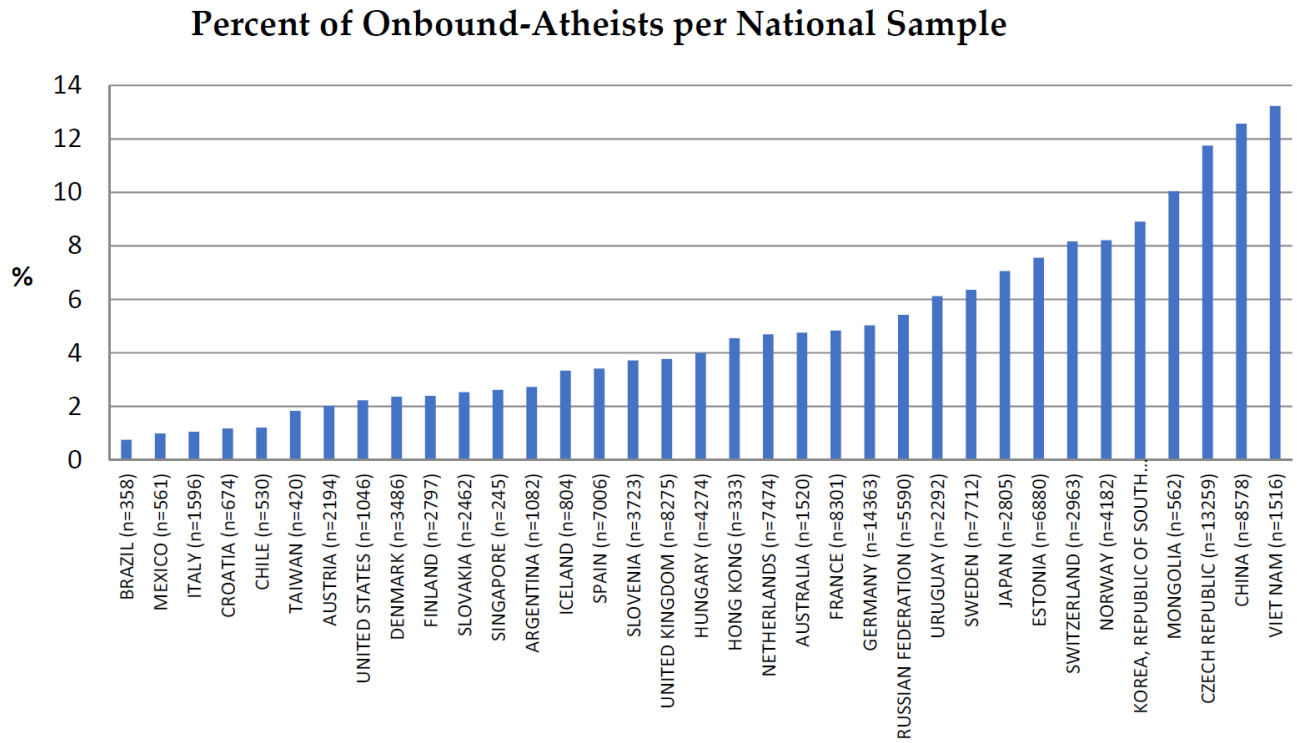


Figure 1. ONBound atheists in percentage of the 35 country samples across all surveys and years.

Source: Bechert, Insa (2012): **Of Pride and Prejudice—A Cross-National Exploration of Atheists' National Pride.** *Religions* , Vol. 12, p. 648 (p.11 of 24)

gesis Leibniz Institute for the Social Sciences

Take Home Message

RTFM!

(to the tiniest detail)

Meet the Experts

Thank you !

gesis Leibniz-Institut für Sozialwissenschaften

Leibniz Gemeinschaft

gesis Leibniz Institute for the Social Sciences

Meet the Experts

"Ex Post-Harmonised" Comparative Survey Data Files at GESIS - and What to Expect Of Them

Markus Quandt & Ivet Solanes-Ros

GESIS Meet-the-Experts Series

June 01 2023

Leibniz Association