Harmonised data for comparative research

Information sheet

Introduction

Comparative social science research requires comparable data across countries, especially data at the level of individual people and households (microdata).

More and more international databases are becoming available to researchers. Some result from cross-national projects adopting ex-ante harmonisation approaches such as developing common concepts, tools and protocols. There have also been significant ex-post harmonisation projects seeking to create comparable datasets by combining multiple existing data sources.

This information sheet details major sources of harmonised data and provides references for those looking to harmonise existing data sources as part of a research project.

> The information sheet accompanies the **CESSDA webinar on *Harmonised Data for Comparative Research*** (DOI: doi.org/10.5281/zenodo.3894145), which included presentations on sources of harmonised data and approaches and tools for harmonising data sources for cross-national research.

Contents

Document info

Status	Public
Author	Jennifer Buckley, University of Manchester
Date	June 2020
Licence	CC-BY 4.0
Version	1
PID	doi.org/10.5281/zenodo.3932361

Sources of harmonised microdata

This section provides information about existing sources of harmonised microdata available for cross-national research. The list here is not intended to be comprehensive but to offer insight into major resources for researchers interested in cross-national social science research.

Cross-national social surveys

There are many cross-national surveys that have the aim of achieving comparable data. Many of these are large collaborative endeavours that have substantial infrastructure to support data collection, processing and dissemination such as the European Social Survey or SHARE – the survey of health, ageing and retirement in Europe.

Cross-national datasets result also from collaborations to develop survey modules to fit into national surveys; this approach is used by the International Social Survey Programme (ISSP) and Comparative study of electoral systems.

List of international surveys

Though many cross-national research projects make data available to researchers, access to data is dispersed across archives and other portals. To help researchers, the CESSDA Data Management Expert Guide (see below for details) includes a list of international surveys that have accessible data.

Box 1: Examples of cross-national survey programmes

European Social Survey
europeansocialsurvey.org
The European Social Survey (ESS) is an academically driven cross-national survey that has been conducted across Europe every two years since 2002. The survey measures the attitudes, beliefs and behaviour patterns of diverse populations in more than thirty nations. The data are available from the project website after registration, for not-for-profit purposes.

SHARE – the survey of health, ageing and retirement in Europe
share-project.org
The Survey of Health, Ageing and Retirement in Europe (SHARE) is a multidisciplinary and cross-national panel database of micro data on health, socio-economic status and social and family networks. Data relates to 123,000 individuals aged 50+ across 27 European countries and Israel. The SHARE data are distributed free to registered users through the SHARE Research Data Center. See the details on the project website.

International Social Survey Programme (ISSP)
w.issp.org
The ISSP is a cross-national collaboration programme to conduct annual surveys on diverse topics relevant to social sciences including social networks, gender roles, religion and citizenship. The program started in 1984 with four founding members - Australia, Germany, Great Britain, and the United States – and has now grown to about 50 member countries from all over the world. The GESIS Data Archive for the Social Sciences create and provide access to the cross-national dataset:
https://www.gesis.org/issp/modules/issp-modules-by-year/

Comparative study of electoral systems (CSES)
cses.org
The CSES is a collaborative program of research among election study teams from around the world. Participating countries include a common module of survey questions in their post-election studies. Since 1996 there have been 5 thematic modules. Registration is required in order to obtain the data via the project website.

Eurostat and microdata from national statistics institutes

ec.europa.eu/eurostat/web/microdata

Eurostat, the statistical office of the European Union, provides access to harmonised microdata for scientific purposes. Box 2 lists Eurostat's collection of microdata. Access is limited to researchers at approved organisations and by application.

For locating information about microdata available from the various national statistics institutes and used by Eurostat, there are two useful tools (which have been compiled by CESSDA data archives):

- MISSY (gesis.org/en/missy/) is an online service platform that provides structured metadata for official statistics. MISSY includes metadata at the study and variable level as well as reports and tools for data handling and analysis. All documentation in MISSY refers to microdata available for scientific purposes.
- CIMES (cimes.casd.eu/) provides overviews of official microdata disseminated for research purposes by the national statistical institutes (NSI) across Europe. It describes data and the procedures for requesting access, information which is otherwise scattered. CIMES does not give direct access to data, only metadata.

Gateway to Global Aging

Gateway to Global Aging (g2aging.org) supports harmonisation across a family of Health and Retirement Studies, which includes SHARE (see above) alongside other such as the U.S. Health and Retirement Study (HRS), the English Longitudinal Study of Aging (ELSA) and Tilda - The Irish Longitudinal Study of Ageing. Gateway to Global Aging maintain an online platform for exploring comparability across the HRS studies and creates harmonised datasets. The harmonised datasets are either accessible via the original studies or via an accessible Stata programme.

Household panel studies and the Cross-national Equivalent File (CNEF)

National household panel studies from across the globe examine comparable topics (see box 3 for national studies). The Cross-national Equivalent File (CNEF), 1970-2017, (cnef.ehe.osu.edu) provides access to a dataset containing a simplified version of the panels for cross-national analysis. There are also guidelines for

Box 2: Eurostat Microdata

European Community Household Panel
European Union Labour Force Survey
Community Innovation Survey
European Union Statistics on Income and Living Conditions
Structure of Earnings Survey
Adult Education Survey
European Road Freight Transport Survey
European Health Interview Survey
Continuing Vocational Training Survey
Community Statistics on information Society
Micro-Moments Dataset
Harmonised European Time Use Survey

Box 3: Household panel studies

Understanding Society
Household Income and Labour Dynamics in Australia (HILDA)
Korea Labor and Income Panel Study (KLIPS)
Panel Study of Income Dynamics (PSID)
Russia Longitudinal Monitoring Survey (RLMS-HSE)
Swiss Household Panel (SHP),
Canadian Survey of Labour and Income Dynamics (SLID)
German Socio-Economic Panel (SOEP)

formulating equivalent variables across countries. Topics in the CNEF include: Demographics, Employment and Income, Health and Psychological well-being. To access the data, potential users need to apply to the data provider in each country for access and then request access to the CNEF via the project website (https://cnef.ehe.osu.edu/apply-for-cnef-data-beta/).

The International Stratification and Mobility File (ISMF)

harryganzeboom.nl/ismf/index.htm

The International Stratification and Mobility File (ISMF) is a project to create a comparative database to study social mobility patterns. The project is notable for its inclusiveness with more than 250 surveys and 56 countries included. Files to create datasets are available via the project website. There are also code/syntax files for creating comparable social stratification measures.

LIS – Luxembourg Income Study

LIS (lisdatacenter.org/our-data) is a long-standing harmonisation project. It collects and harmonises socio-economic micro datasets from upper- and middle-income countries into databases relating to wealth and income. Due to confidentiality restrictions relating to the underlying microdata, access is via a remote service (LISSY), which is available to registered researchers and exclusively for non-commercial purposes. There is also an online table maker.

Survey Data Recycling (SDR) project

asc.ohio-state.edu/dataharmonization

The SDR project is creating a multi-country multi-year database with a thematic focus on political engagement, trust and social capital. The core of the SDR database is information pooled from 24 international survey projects: with 3733 national samples of 187 project-waves. The SDR database is available via the Harvard Dataverse: https://doi.org/10.7910/DVN/VWGF5Q.

Time-use studies

There have been projects to harmonise time-use studies that use a comparable design through both

- HETUS (a Eurostat project): ec.europa.eu/eurostat/web/microdata/time-use-survey
- Multinational Time-use Study: timeuse.org/mtus

Harmonising data sources: selected resources

Beyond existing harmonised data sources, we can create new harmonised datasets for comparative research. The previous section includes examples of ex-post harmonisation projects, where datasets have been compiled using multiple individual surveys. This final section lists some resources and reading related to ex-post harmonisation.

CESSDA webinar on Harmonised Data for Comparative Research
(DOI: 10.5281/zenodo.3894146)
The webinar discussed the sources of harmonised data here alongside the following presentations on harmonisation:

- Ex-post harmonization in the Survey Data Recycling, SDR, Project (Irina Tomescu-Dubrow)
- Harmonization and cross-national comparative research (Dr Kristi Winters, GESIS)

Guidelines for Best Practice in Cross-Cultural Surveys (CCSG)
ccsg.isr.umich.edu/
The CCSG guidelines were developed as part of the Comparative Survey Design and Implementation (CSDI) Guidelines Initiative. The aim is to promote guidelines and best practice for the conduct of multinational, multicultural, or multiregional surveys, referred to as "3MC" surveys. The intended audience is researchers and survey practitioners planning or engaged in comparative survey research across cultures or countries. The guideline includes a chapter on the topic of harmonisation, with a small section on ex-post harmonisation.

Survey Data Recycling (SDR) project
asc.ohio-state.edu/dataharmonization
The SDR project is developing a methodology for harmonising ex-post data pooled from cross-national survey projects in the social sciences. One development is to create harmonisation control variables to help account for error from the source survey and from harmonisation. The project website includes numerous papers and presentations, including content from a workshop on Building Multi-Source Databases for Comparative Analyses. They also produce a Newsletter on Survey Data Harmonization in the Social Sciences (asc.ohio-state.edu/dataharmonization/newsletter).

Advances in Comparative Survey Methodology
Johnson, T.P., B-E Pennell, I. A. L. Stoop, & B. Dorer (eds). 2018. *Advances in Comparative Survey Methodology: Multinational, Multiregional and Multicultural Contexts (3MC)*, Wiley Hoboken, New Jersey

This book examines recent advances in methodology and operations as well as the technical developments in international survey research. There are several chapters dedicated to ex-post harmonisation.

CESSDA Data Management Expert Guide
cessda.eu/DMGuide

Data harmonisation work requires effective data management. The CESSDA Data Management guide covers the whole research data lifecycle from planning, organising, documenting, processing, storing and protecting data to sharing and publishing.

CharmStats
gesis.org/en/services/data-analysis/data-harmonization/charmstats

CharmStats is a software package developed by GESIS that allows you to work with your own variables, to document the process and electronically publish your completed harmonization work for review and citation.

Synergies for Europe's Research Infrastructures in the Social Sciences (SERISS)
seriss.eu/training/training-overview

Synergies for Europe's Research Infrastructures in the Social Sciences (SERISS) was a four-year project that aimed to strengthen and harmonise social science research across Europe. Part of the project was to provides information and resources to those who are interested or working in cross-national social science research. Several tools (https://seriss.eu/training/tools/) have been developed relating to data management, data handling and harmonisation, as well as on statistical analysis to a wider social science community. The work included the surveycodings.org project, see below for details.

Surveycodings.org

SurveyCodings.org is a free service for survey projects, data archives and researchers measuring and coding socio-economic background variables. It provides questionnaires, data collection tools, coding frames and datasets based on standard statistical classifications for many countries and languages for the following variables: industry, occupation, educational attainment, employment status, socio-economic status, and social networks. The materials can be used to collect harmonised and cross-nationally comparable data as well as for post-hoc harmonisation of already collected data (e.g. coding of open-ended questions).