



Vulnerability Indicators in European Datasets -

Harmonizing and integrating data on vulnerable groups

Janna Besamusca and Stephanie Steinmetz (University of Amsterdam)

InGRID2 Data Forum: Data Pooling - Opportunities and Challenges, CELSI Bratislava, 11. October. 2019





The InGRID project in a nutshell http://www.inclusivegrowth.eu/







About InGRID2 - Objectives

- Integrate and innovate existing European Social Sciences
 Research Infrastructures on
 - Poverty and living conditions
 - Working conditions and vulnerability
- by **improving**:
 - Transnational data access
 - Organising mutual knowledge exchange activities
 - Improving methods and tools for comparative research
- In order to create new & better opportunities for development of evidence-based European policies on Inclusive Growth



About InGRID2 - Organization

18 partners in a consortium Various types of activities:

- Summer schools
- Expert workshops and roundtables
- Data forums
- Visiting grants to data infrastructures (Call 8 October 2019: visits between 03/2020 and 08/2020)
- E-portal
- Virtual data access

Joint research activities



About InGRID2 – Organization II

Poverty & Living Conditions

Data integration & harmonisation (situation & policy)

Evaluation & analysis tools (simulation)

Indicator-building (visualisation, policy innovation)

Working Conditions & Vulnerability

Data integration & harmonisation (situation & policy)

Evaluation & analysis tools (simulation)

Indicator-building (visualisation, policy innovation)



Motivation

- Researchers are confronted with an overwhelming amount of surveys.
- Interest in particular populations (such as refugees, sexual/gender minorities etc.) ⇒ search, access, explore, harmonize data ⇒ TIME CONSUMING!
- Lack of an inventory providing guidance / overview which supports researchers in this endeavour.
- Lack of multidimensional indicators regarding working conditions of vulnerable groups.



Some already existing initatives





👫 Jobportale in der Schweiz Onli 🗙	🍕 Jobs, Stellen & Stellenangebot: X 🛛 🔲 Exploring the dataset of survey 🛛 🖉	🔀 Data Harmonization - Cross-Cu X CCSG_Full_Guidelines_2016_Version X Microsoft Word - Data Harmonizati X 🕇		- o ×
$\left(\leftarrow ight) ightarrow$ C' $ m (a)$	🛈 🛈 🏠 https://ccsg.isr. umich.edu /index.php/chapters/data-harmoni	zation-chapter	E V &	⊻ II\ 🗊 📽 Ξ
	CRO Home	SS-CULTURAL SURVEY GUIDELINES		
	💡 You Are Here 🍗 Chapter	s > Data Harmonization		
	Study Design and	DATA HARMONIZATION		
	Organizational Structure			
	Tenders Bids & Contracts	<u>A</u> <u>A</u>		
	Sample Design	Peter Granda and Emily Blasczyk, 201 6		
	Questionnaire Design	Introduction		
	Instrument Technical Design	Guidelines		
	Translation	1. Decide what type of harmonization strategy to employ, taking into account that many		
	Overview	harmonization efforts will require some combination of strategies.		
	Management and Budgeting	2. When deciding which variables to harmonize, create an initial plan and define clear objectives		
	Team	about what you want to achieve. The plan should include making all data conversions		
	Scheduling	3. Focus on both the variable and survey levels in the harmonization process.		
	Shared Language Harmonization	4. Develop criteria for measuring the quality of the harmonization process. This includes testing it		
	Assessment	with users knowledgeable about the characteristics of the underlying surveys, the meaning of		
	Tools	source variables, and the transformation of source variables into target variables.		
	Adaptation	5. Provide the widest range possible of data and documentation products about the entire		
	Pretesting	Further Reading		
	Interviewer Recruitment,			
	Selection, and Training			
	Data Collection	Introduction		
	General Considerations	Harmonization refers to all efforts that standardize inputs and outputs in multinational, multicultural, or		
	Face-to-Face Surveys	multiregional surveys, which we refer to as "3MC" surveys.		
	Telephone Surveys	Harmonization is a generic term for procedures used predominantly in official statistics that aim at		
	Self-Administered Surveys	achieving, or at least improving, the comparability of different surveys and measures collected. The term is		
	Paradata and Other Auxiliary	closely related to that of standardization (see Sample Design and Questionnaire Design). Harmonizing		
1	Data	procedures may be applied in any part of the survey lifecycle, such as study design, choice of indicators,		
	Data Harmonization	question wording, translation, adaptation, questionnaire design, sampling, data collection, data coding,		14.67
\mathcal{P} Typ hier om te zoeker	n 🗏 🔁 🗖 🛱 🕑 👯	<u> </u>		^ [↑] ♥ ■ <i>@</i> 10-10-2019



Objective

- To provide
 - Overview of current ability of European datasets to identify nine vulnerable groups and of consequences of different identification strategies.
 - Harmonization proposals to allow researchers to pool data on otherwise underrepresented vulnerable groups.
 - Recommendations regarding most effective instruments for the identification and measurement of vulnerable groups.

Development of multidimensional indirators for vulnerable

Harmonising and integrating data on vulnerable groups by creating an inventory how they can be identified and harmonised across 28 EU-wide mirco-level data sets



Selection criteria for datasets

- Must cover multiple countries in the European area
- Must measure concepts related to the labour market and to demographic characteristics
- Must be recently conducted and/or part of an ongoing data collection (exception: ECHP)
- May be cross-sectional or longitudinal
- May cover any topics in the social realm (e.g. incomes, attitudes, working conditions)



Datasets

- EU surveys (EU-SILC, ECHP, EU LFS, Structure of earning Survey, Adult Education Survey, European Health Survey, European Community statistics on information society, European Household Budget Survey, Eurobarometer)
- Eurofound surveys (ECS, EWCS, EQLS)
- Wage and Salary surveys (WageIndicator Salary Survey & Salary checker, Luxembourg Income Study, Industry Wages)
- Demographic surveys (IEPM, GGP, SHARE)
- Time Use surveys (MTUS)
- Social and Attitudinal surveys (ESS, ISSP, EVS, WVS)
- Skills and Education surveys (PIACC, PISA)



Challenges to survey hard-to-survey populations

Hard-to-survey populations are those that create problems for one or more key survey operations (Tourangeau et al. 2014)

- Hard-to-sample (rare populations with no population-specific frame)
- Hard-to-identify (based on hidden or stigmatizing characteristics)
- Hard-to-find or contact (mobile populations; pop's with access issues)
- -Hard-to-persuade (resistant; alienated)
- Hard-to-interview (language barriers)



ерітер ву Roger Tourangeau, Brad Edwards, Timothy P. Johnson, Kirk M. Wolter, and Nancy Bates



Selection of hard-to-survey populations

- In the framework of work and labour markets
 - Migrants / ethnic groups (undocumented)
 - Religious groups
 - Sexual / gender minorities
 - Disabled

⇒ Becomes more challenging when intersections are considered with more common socio-demographic characteristics like gender, age, education etc. (female refugees, etc.)



Selected populations

Hard-to-reach population	Operationalisation
Gender identity	Man / Woman and potentially other measurements
Migrant background	 Identified through a series of questions: Country of residence / work / education / birth Mother/father country of birth Year of arrival in country of survey Reason for coming to country of survey
Nationality	Citizenship status
Ethnicity	Self-identified membership of ethnic group
Disability	Unable to work due to long-term illness or disability
Religion*	Self-identified membership of religious denominations and religiosity indicators
Sexual orientation*	"Which of the options best describes how you think of yourself?" (Heterosexual/Straight, Gay/Lesbian, Bisexual, Other). (country coverage: Belgium, Italy, Netherlands, Portugal, Spain, UK)



Our approach: mapping measurements

Step 1: mapping measurement of each of these hard-to-survey populations across the datasets

- Select most current year and harmonized variables
- Collect meta-data (country coverage, years, interview mode, etc)
- Collect relevant questions and answer categories

Step 2: exploring consequences of sampling, survey mode and operationzalization of measurement

- On distribution
- On non-response
- On drop-out



Inventory example: metadata

#	Dataset	Years	Sample size	Region	Publisher	Sampling frame	probability sample?	Mode of interviews
1	WageIndicator full survey	2000-2017		world	WIF	volunteer	n	web-based self-interview
19	ISSP	1986-2016		World	ISSP	population 18+	у	mixed
20	European Values Survey	1981-2008*		Europe	EVS	population 18+	V	face to face
21	World Values Survey	1981-2014		world	WVSA	population 18+	v	mainly f2f, mixed
22	European social survey	2002-2016		Europe	ESS - ERIC	population 15+	V	face to face
23	PIACC survey of adult skills	2008-2019	5000/ country	OECD	OECD	age 16-65	у	CAPI



Inventory example: variable data

		22_European Social Survey
Variable	Value	Label
brncntr	question	Were you born in [country]?
		11 Yes
		22 No
		77 Refusal
		88 Don't know
		99 No answer
cntbrthc	API	In which country were you born?
livecnta	question	What year did you first come to live in [country]?
		year
facntr	question	Was your father born in [country]?
		1 1 Yes
		22 No
		77 Refusal
		88 Don't know
		99 No answer
fbrncntb	API	In which country was your father born?
mocntr	question	Was your mother born in [country]?
		1 1 Yes
		22 No
		77 Refusal
		88 Don't know
		99 No answer
mbrncntb	API	In which country was your mother born?



Preliminary overview

Vulnerable group	Operationalisation
Gender identity	Generally measured by the same, often interviewer-coded binary variable . Included in all surveys.
Sexual orientation*	Mostly derived from household grid . WageIndicator test question asks for orientation, ESS and Eurofound ask for experiences of discrimination based on sexual orientation, GGP ask for sex of non-resident spouse
Migrant background	Most surveys ask which for country of birth. Difference in inclusion of the same question for mother and father of respondent as well as for detail of the included list of countries of birth.
Nationality	Citizenship status is rarely included. Sampling frames differ in the inclusion or exclusion of non-citizens.
Ethnicity	Included in selected surveys (mainly Social and Values surveys) and much variation in the available detail of ethnic groups.
Disability	Large variation in measurement. Includes self-identification, being out of work due to disability, being hampered in execution of work due to temporary or permanent disability or sickness, limited in daily activities, discriminated against based on disability.
Religion	WageIndicator, Social and Values surveys include self-identification items, religiosity measures and discriminated against (latter also in Eurofound surveys)



Example: Disability

Disabled persons suffer from both employment and wage gaps (linked to labour market discrimination) (Kruse et al, 2018; Mann & Wittenburg, 2015)

Inventory shows \Rightarrow datasets rely on wildly different identification strategies to measure disability status. Most common instruments are:

- People who are unable to work (main economic status (WI, AES, EHIS, HBS, ISSP)
- Individuals'/households' income from sickness or invalidity benefits (ECHP, SHARE)
- Sickness, injury or incapacity as motive for not seeking paid employment or further education or training (ECHP, AES)
- Respondents' self-reported chronic mental health problems or physical disabilities (ECHP, EHIS, EWCS, EQLS, GGP)
- Respondents' self-reported limitations in daily or work activities (ECHP, EHIS, EWCS, EQLS, ESS, SHARE)
- Respondents' self-reported experiences of discrimination based on their disability (EWCS, ISSP, ESS)
- Respondents' self-reported inability to continue performing their job at all or at the previous intensity (SHARE)



Proportion of sample reporting a disability







Concluding remarks and next steps

- Identification of disable people in European datasets is highly dependent on chosen instrument.
- Next steps:
 - examining effect of different questions on item nonresponse, group size and relation between disability and labour market vulnerability.
 - comparing with non-probability samples to evaluate in how far they allow for a better incusion of such populations
- Challenges: how to address common exclusion of institutionalized population, and regular grouping together of disability and illness



THANK YOU

Co-ordinator Monique Ramioul



RESEARCH INSTITUTE FOR WORK AND SOCIETY

Partners

TÁRKI Social Research Institute Inc. (HU) Amsterdam Institute for Advanced Labour Studies – AIAS, University of Amsterdam (NL) Swedish Institute for Social Research - SOFI, Stockholm University (SE) Economic and Social Statistics Department, Trier University (DE) Centre for Demographic Studies - CED, University Autonoma of Barcelona (ES) Luxembourg Institute of Socio-Economic Research - LISER (LU) Herman Deleeck Centre for Social Policy – CSB, University of Antwerp (BE) Institute for Social and Economic Research - ISER, University of Essex (UK) German Institute for Economic Research - DIW (DE) Centre for Employment and Work Studies - CEET, National Conservatory of Arts and Crafts (FR) Centre for European Policy Studies - CEPS (BE) Department of Economics and Management, University of Pisa (IT) Department of Social Statistics and Demography - SOTON, University of Southampton (UK) Luxembourg Income Study - LIS, asbl (LU) School of Social Sciences, University of Manchester (UK) Central European Labour Studies Institute – CELSI (SK) Panteion University of Social and Political Sciences (GR) Central Institute for Labour Protection – CIOP, National Research Institute (PL)

j.besamusca@uva.nl s.m.steinmetz@uva.nl

Contact:

InGRID-2

Integrating Research Infrastructure for European expertise on Inclusive Growth from data to policy Contract N° 730998

For further information about the InGRID-2 project, please contact inclusive.growth@kuleuven.be www.inclusivegrowth.eu p/a HIVA – Research Institute for Work and Society Parkstraat 47 box 5300 3000 Leuven Belgium