

Variable Harmonization in the Social Sciences

CESSDA Training Group,
2019

*Dr. Kristi Winters, GESIS Leibniz Institute for the Social Sciences
Mr. Janek Bruker, GESIS Leibniz Institute for the Social Sciences*

GESIS, Köln, August 2019

 [cessda.eu](https://www.cessda.eu)

 @CESSDA_Data



WHY?



Why Harmonizing?

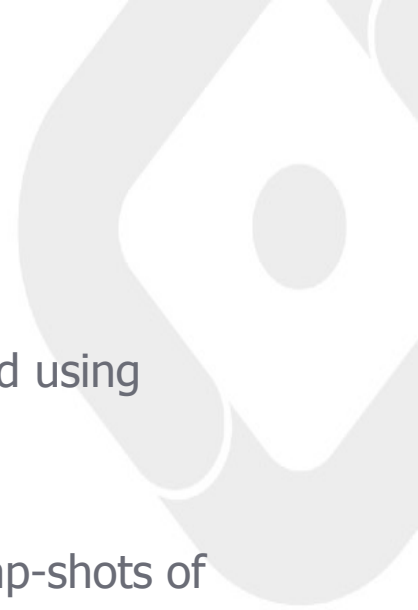
- Necessary for cross-national comparative research
- Contributes to basic (empiric) knowledge
- Allows to conclude impact of policies on societies and human behaviour
- Enables to compare impacts across nations, cultures or over time



Why harmonizing?

Besides research questions a lot of policy questions can be answered using comparative (harmonized) data

- *What* is happening to outcomes of interest?
 - Cross-sectional samples of population of interest = regular snap-shots of (social and economic) indicators
 - Panel data provides information about the evolution of these indicators over time, possibly even over the complete lifecycle of a respondent



Why harmonizing?

- *Why* are given outcomes observed?
 - Disentangle the impact of (micro or macro) characteristics or policy decisions on certain outcomes, e.g. minimum wage on economic well-being
- *How* can “bad” outcomes be reduced and “good” outcomes encouraged?
 - Structural model: investigating in detail structure of decision leading to an outcome of interest
 - Natural experiment: impact of changes in policies (e.g. by comparing countries with different policies)



Why harmonizing?

The information **researchers** gain from harmonized and thus comparative data can be provided to **policy-makers** to see the impact of their policies

→ observe outcome, develop new or improve running policies

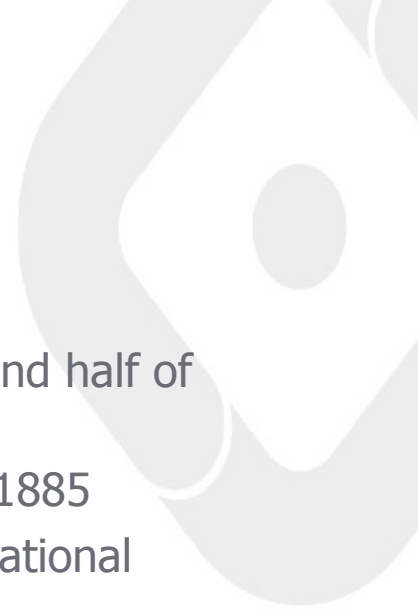


HISTORY OF DATA HARMONIZATION



Origin

- Developed with general approaches to statistical research in second half of 19th century
- Establishment of *International Statistical Institute* in The Hague, 1885
 - One main activity: standardization and harmonization of international statistics



Driven by research and politics

- International organizations and institutions began collection of international data at beginning of 20th century
 - E.g. International Labour Office (1901), International Agricultural Institute (1905)
- Discrepancies across countries demanded approximation → *Harmonization/Standardization*
 - Establishment of Statistical Office of the United Nations, Conference of European Statisticians

Driven by research and politics

- Increasing demand of cross-national and cross-temporal comparable data
 - On one hand: research interest, e.g. analysis of social change
 - On other hand: governments and (international) administrative bodies for improvement of policy making process, e.g. effectiveness of minimum wage

International surveys

- Increasingly surveys began to react to demands and have employed international standards to produce comparable data
 - E.g. ISSP, ESS, EVS
 - In 2001, ISSP changed strategy for background variables from ex-post to ex-ante output (see later) to guarantee more comparability



CONCEPTUAL ASPECTS OF DATA HARMONIZATION



Overview

- What is harmonization?
- Harmonization strategies
- Harmonization processing (incl. examples)
- International Standards
- Quality checks



What is harmonization?

- **Standardization** of inputs and outputs in comparative statistical analysis (Granda & Blasczyk, 2016)
- “(...) creation of a desired degree of **comparability** between statistics of different countries.” (Ehling, 2003: 17)
- “(...) generic term for procedures aimed at achieving or at least improving the **comparability** of different surveys.” (Granda et al., 2010)

What is harmonization?

Goal: "(...) create data that measure the same conceptual variable and that are measured in the same units" (Burkhauser et al., 2005)

→ Resulting: single integrated dataset for comparative statistical analysis, either including all variables to serve as general resource or specific variables to serve as study resource

What is harmonization?



- Central element: **Comparability**
- Comparison can be *spatial* (cross-national) or *temporal* (longitudinal) comparability
- First step: universal reference concept
 - How to operationalize and make measurable
 - Find universal definition, classification, etc. for adequate categorization of all particularities in participating countries or cohorts

What is harmonization?

- Generally every part of the survey lifecycle can be harmonized
 - E.g. question wordings, choice of indicators, field works, questionnaire design
- Different harmonization strategies offer approaches mainly different regarding the time point and extent of harmonization



Harmonization strategies



Granda and Blasczyk, 2016; Granda et al., 2010:

- Input vs. Output
- Ex-ante vs. ex-post

Doiron et al., 2012; Fortier et al., 2011:

- Prospective vs. Retrospective
- Stringent vs. flexible
- Both classification systems are very similar

Harmonization strategies

Input/prospective	Output/retrospective
Ex-ante per definition	Ex-ante or ex-post
Stringent or flexible	Flexible per definition

Keep in mind: Categorization presents ideal types that de facto often overlap

Harmonization strategies



(Ex-ante) Input Harmonization

Following strict standards from the beginning of the survey process and applying identical survey procedures and questionnaires in each survey

Ex-ante Output Harmonization

Considering comparability in survey planning and implementing comparable measurements in data collection process “to capture the same concept” among different surveys (e.g. exact wording or some answer categories can vary)

Ex-post Output Harmonization

Using items from different existing surveys that were not designed to be comparable but that are edited for comparative analysis

Harmonization strategies



(Ex-ante) Input Harmonization

Following strict standards from the beginning of the survey process and applying identical survey procedures and questionnaires in each survey

- **Pro**: long-term standardization; can lead to more coherence if meeting national data requirements; usually best for international comparability
- **Con**: initially higher costs; loss of flexibility

Harmonization strategies

Ex-ante Output Harmonization

Considering comparability in survey planning and implementing comparable measurements in data collection process “to capture the same concept” among different surveys (i.e. exact wording or some answer categories can vary)

- **Pro:** compromise btw. International standards and national requirements; relatively low cost
- **Con:** conflict btw. International and national goals and quality criteria

Harmonization strategies



Ex-post Output Harmonization

Using items from different existing surveys that were not designed to be comparable but that are edited for comparative analysis

- **Pro:** Transitional solution; re-use of existing data
- **Con:** usually lowest international comparability; possible cost-efficiency

Harmonization strategies

Distinction by time point of data collection:

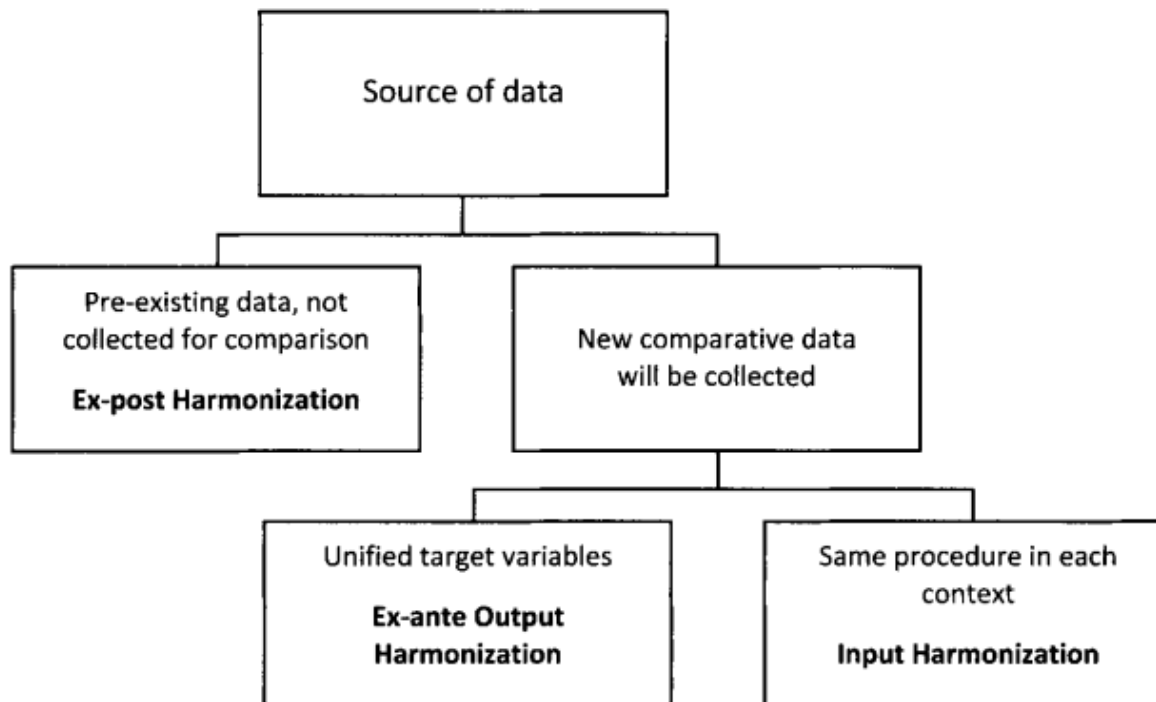
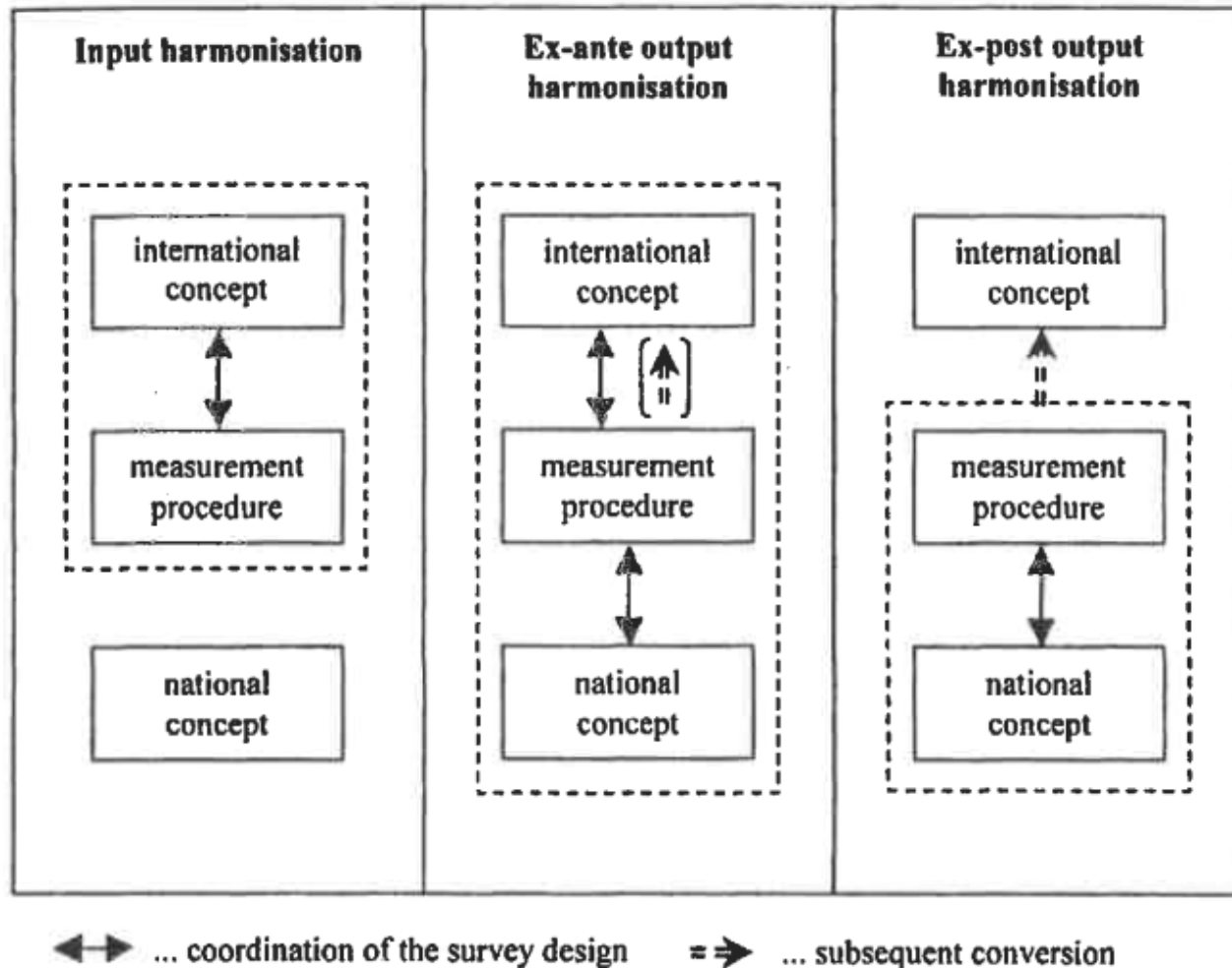


Figure 17.1. Three Approaches to Harmonization

Harmonization strategies



„Harmonisation strategies“ in Ehling, M. (2003). Harmonising data in official statistics. In Hoffmeyer-Zlotnik and Wolf (eds.), *Advances in cross-national comparison* (p. 23). New York, Kluwer Academic/Plenum Publishers.

Harmonization strategies



Stringent-prospective Harmonization

Group of studies decides to apply identical data collection tools and procedures

Flexible-prospective Harmonization

Agreeing on common variables but leaving space for individual studies to adapt data collection to their specific (e.g. cultural or regional) needs

Flexible-retrospective Harmonization

Using pre-assessed compatible data from several existing studies to pool a set of target variable by applying formal harmonization rules (conversion process)

Harmonization strategies



Stringent-prospective Harmonization

Group of studies decides to apply identical data collection tools and procedures

- **Pro:** Straightforward harmonization with standardized collection of data; guarantees compatibility across data
 - **Con:** No adaption to specific contexts of individual studies during data collection; inserting standard measures across multiple studies can get very challenging
- Very strict *input* and thus always *ex-ante*

Harmonization strategies

Flexible-prospective Harmonization

Agreeing on common variables but leaving space for individual studies to adapt data collection to their specific (e.g. cultural or regional) needs

- **Pro:** better in measuring regional/cultural/etc. variation; respects center-specific scientific foci
 - **Con:** “rigorous assessment of heterogeneity is required” (Doiron et al., 2012); data need to be assessed as similar enough for specific scientific purposes
- Mainly *ex-ante output*, might include *input*

Harmonization strategies

Flexible-retrospective Harmonization

Using pre-assessed compatible data from several existing studies to pool a set of target variable by applying formal harmonization rules (conversion process)

- **Pro:** re-use of data for secondary analysis; relatively modest time and costs
- **Con:** appropriate expertise and adequate methodologies needed; limited quantity of compatible data

➤ *Always ex-post output*

Harmonization strategies

Distinction by level of standardization:

Ex-post Output,
Flexible-
retrospective

Ex-ante Output,
Flexible-
prospective

Input,
Stringent-
prospective

→ Complete
standardization

Harmonization strategies


- To decide on strategy of data harmonization it is easier to follow a **variable-specific approach** than a survey oriented one
- European Social Survey (ESS) includes some *ex-ante output* and some *input* harmonized variables = *flexible-prospective* harmonization approach
- Some variables follow a strictly standardized pattern while others depend on country-specific conditions

Harmonization Processing



- Decide on your strategy as soon as possible to include data harmonization into the survey lifecycle when applicable
- Planning is everything!
- Create a data (harmonization) processing plan

Data Harmonization Processing Plan

- Follow a systematic design
 - Include experts and methodology groups
 - Include country-specific experts if necessary
 - Gather feedback on the analytic usefulness of the data
 - Implement (external) quality controls
 - Use software facilitating variable comparison and harmonization planning
- 

Harmonization Processing

Rules for harmonization (see Hoffmeyer-Zlotnik, 2008: 11f.; Hoffmeyer-Zlotnik & Wolf, 2003: 405)

- Agree on common definition of each variable that is measured
- Common definition refers to comparable elements of participating countries
- Analysis of national concepts and structures by country specialist
- Identify areas of common ground underlying national concepts and structures

Harmonization Processing

- Decide on valid indicator respecting variable of interest and national manifestations
- Choose harmonization strategy
- Test survey/classification instrument with regard to empirical structures in country and correspondence to logic of joint definition
- Ensure that average respondents from different countries understand survey instrument in intended manner and are able to answer question

Harmonization Processing



A crucial step in ex-ante harmonization is the creation of a cross-cultural survey with the highest degree of comparability possible...

Harmonization Processing

Referring to list of Hurst, B. C. & Patrick, D. I. (1998), Harkness, J. et al. (2003) name the following steps for cross-cultural questionnaire design:

- (1) Review of literature and existing instruments
- (2) Establishment of a conceptual framework
- (3) Elicitation of items
- (4) Evaluation of cross-cultural equivalence

Harmonization Processing



- (5) Development and refinement of draft questionnaire
- (6) Evaluation of psychometric properties
- (7) Evaluation of responsiveness
- (8) Preparation of users' manual and, if relevant, a scoring scheme
- (9) Submission to a supervisory council (or other signing-off procedure) and distribution

Excursus: Translation approaches

With existing materials

Adopting

- Translate (usually mono-cultural) survey items
 - Translation team

Adapting

- Modifying existing items with respect to cultural appropriateness
 - Expert team with country- and survey-specific knowledge



Excursus: Translation approaches

Without existing materials

Sequential

- Source questionnaire developed and tested in context and language of development and translated afterwards
 - (Small) expert group

Parallel

- Includes cross-cultural input from target cultures and languages during development process of source questionnaire
 - Multicultural expert team

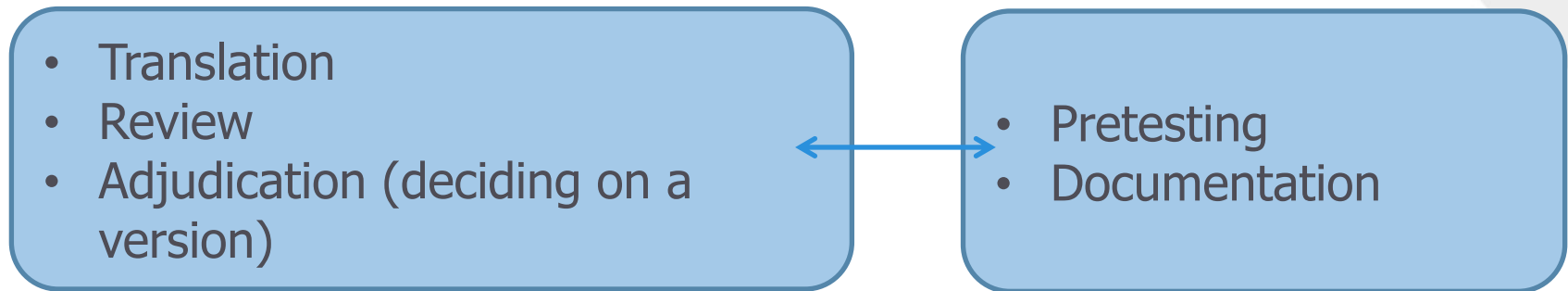
Simultaneous

- Producing different language versions at the same time, can become a mixture of country-specific and cross-country questions
 - Multicultural expert team



Excursus: Translation approaches

5 basic steps in producing final multilingual questionnaire versions:



This usually requires three different groups of staff:

- Translators
- Translation reviewers
- Translation adjudicators

Harmonization Processing



Ex-post harmonization steps (Granda et al., 2010)

- (1) identify useable (existing) data
- (2) define variables of interest
- (3) databank with target of interest as cases and corresponding variables from selected survey
- (4) decide on mapping schemes
- (5) recode original variables to target

Harmonization Processing

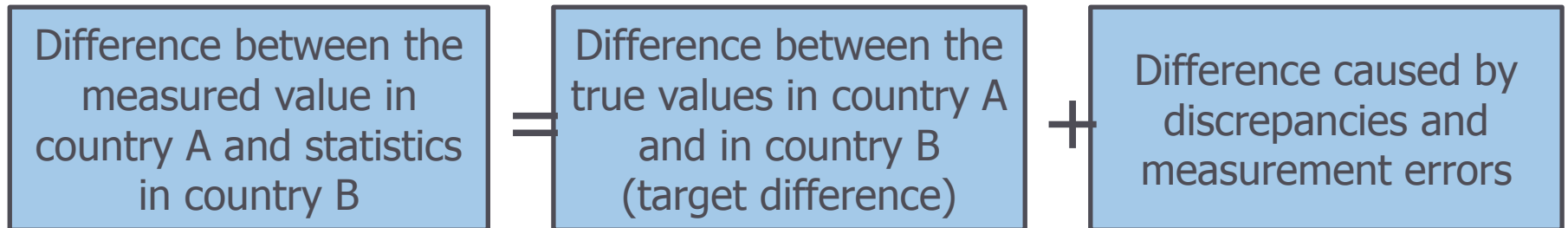


Keep in mind during process:

- Multi-/mono-cultural nature of (cross-) national surveys (“respect context”)
- File-level attributes
 - Sample weights and designs, variance estimation
- Harmonization decisions should be made variable-specifically
 - Varying strategies can be adopted in same survey

Challenges due to country differences

- Regarding a difference between two country values, a researcher must basically distinguish between two components:

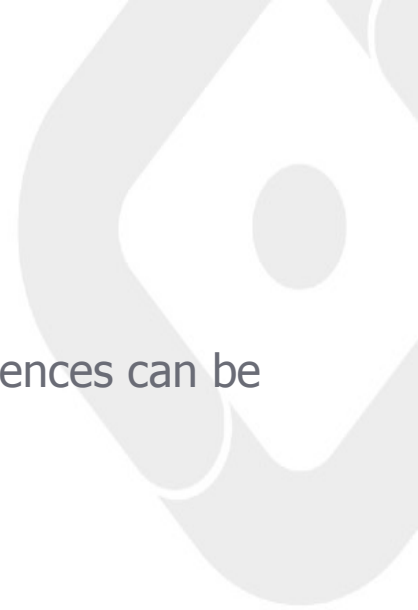


Graphic in Ehling, M. (2003). Harmonising data in official statistics. In Hoffmeyer-Zlotnik and Wolf (eds.), *Advances in cross-national comparison* (p. 24). New York, Kluwer Academic/Plenum Publishers.

Challenges due to country differences

Thus, apart from the target difference that we want to find out, differences can be caused by:

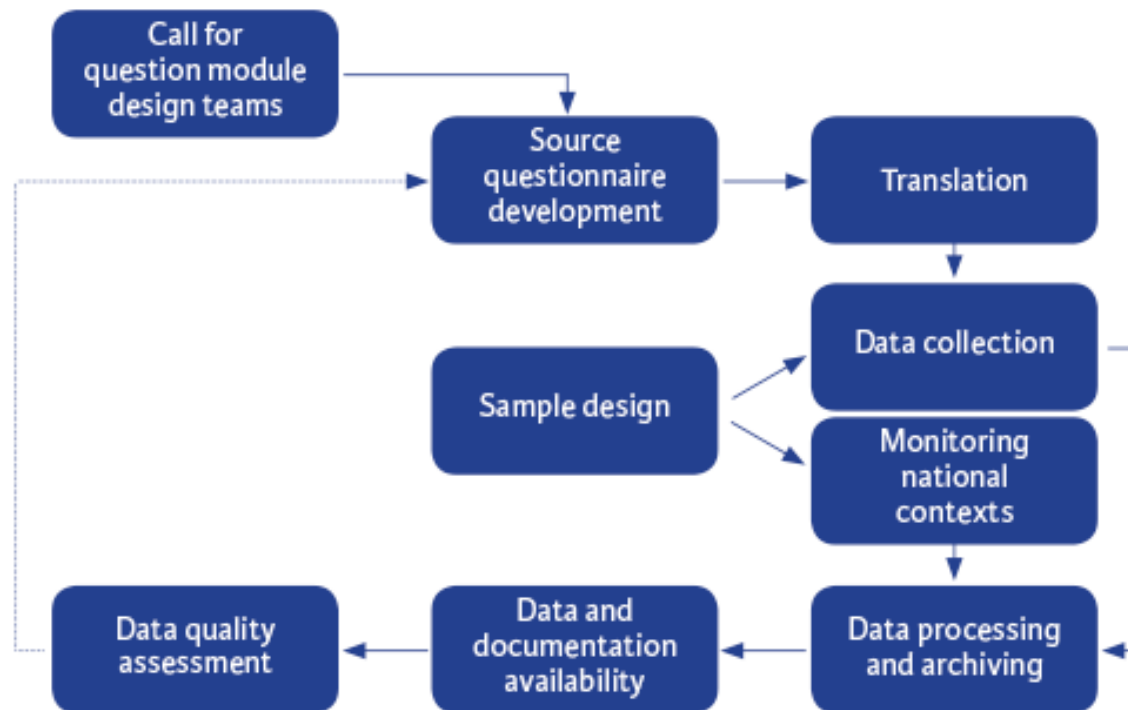
- Differences in **scales** of measurement or coding schemes
- Differences in **definition** of concepts or variables
 - E.g. concept of citizenship
- Differences in **sampling** designs and **collection** methods (survey level comparability)
- Changing **legal** environment around data protection and data privacy (e.g. GDPR)



ESS: Harmonization process

- Standardized set of survey specifications for each participating national coordinator
- Key elements: (1) questionnaires and translation; (2) preparation of data collection and sampling; (3) data collection; (4) data processing and delivery
- For Translation: Translation Team, Translation Expert Panel

ESS: Harmonization process



EVS: Harmonization process

- Harmonization including input and ex-ante output elements
- Careful translation process on basis of Master questionnaire

<https://europeanvaluesstudy.eu/methodology-data-documentation/survey-2017/methodology/>

EVS: Translation process

- Country teams following a team approach according to the TRAPD method
<https://europeanvaluesstudy.eu/methodology-data-documentation/survey-2017/methodology/the-trapd-method-for-survey-translation/>
- Process assisted by web-based platform: Translation Management Tool (TMT)
<https://europeanvaluesstudy.eu/methodology-data-documentation/survey-2017/methodology/the-translation-management-tool-tmt/>
- Cooperation between shared-language countries (e.g. Russian, French, German)
<https://europeanvaluesstudy.eu/methodology-data-documentation/survey-2017/methodology/translation-harmonization-between-shared-language-countries/>

ESS: Input harmonized

Source:

C2 CARD 20 Using this card, how often do you meet socially³¹ with friends, relatives or work colleagues?

Never	01
Less than once a month	02
Once a month	03
Several times a month	04
Once a week	05
Several times a week	06
Every day	07
(Refusal)	77
(Don't know)	88

Germany:

SCLMEET	Wie oft treffen Sie sich mit Freunden, Verwandten oder privat mit Arbeitskollegen? Bitte benutzen Sie Liste 20.	
C2		
	Nie	1
	Weniger als einmal im Monat	2
	Einmal im Monat	3
	Mehrmals im Monat	4
	Einmal in der Woche	5
	Mehrmals in der Woche	6
	Täglich	7
	(Weiß nicht)	88

Switzerland:

C2 • KARTE 20. Benutzen Sie diese Karte und sagen Sie mir, wie häufig treffen Sie sich mit Freunden, Verwandten, Bekannten oder privat mit Arbeitskollegen?

1	Nie
2	Weniger als einmal pro Monat
3	Einmal pro Monat
4	Mehrmals pro Monat
5	Einmal pro Woche
6	Mehrmals pro Woche
7	Jeden Tag
77	(Antwort verweigert)
88	9 (Weiss nicht)

France:

[S] C2 - Dites-moi s'il vous arrive, en dehors du travail, de fréquenter des amis, des membres de votre famille ou des collègues ?

Veuillez utiliser cette carte pour répondre.

ENQUÊTEUR : MONTRER LISTE 20

- 01) Jamais
- 02) Moins d'une fois par mois
- 03) Une fois par mois
- 04) Plusieurs fois par mois
- 05) Une fois par semaine
- 06) Plusieurs fois par semaine
- 07) Tous les jours
- 08) [Refus]
- 09) [Ne sait pas]

ESS: Ex-ante output harmonized

- From round 5 on a new variable measuring education has been developed: EDULVLB
 - 3-digit hierarchical coding framework
- Dr. Silke Schneider (GESIS) participated as an expert in the development process
- The variable is created on the basis of country-specific measures

ESS: Ex-ante output harmonized

3.1 Highest level of education, EDULVLB

Code	Label
0	Not completed ISCED level 1
113	ISCED 1, completed primary education
129	Vocational ISCED 2C < 2 years, no access ISCED 3
212	General/pre-vocational ISCED 2A/2B, access ISCED 3 vocational
213	General ISCED 2A, access ISCED 3A general/all 3
221	Vocational ISCED 2C >= 2 years, no access ISCED 3
222	Vocational ISCED 2A/2B, access ISCED 3 vocational
223	Vocational ISCED 2, access ISCED 3 general/all
229	Vocational ISCED 3C < 2 years, no access ISCED 5
311	General ISCED 3 >=2 years, no access ISCED 5 *
312	General ISCED 3A/3B, access ISCED 5B/lower tier 5A
313	General ISCED 3A, access upper tier ISCED 5A/all 5
321	Vocational ISCED 3C >= 2 years, no access ISCED 5
322	Vocational ISCED 3A/3B, access 5B/lower tier 5A
323	Vocational ISCED 3A, access upper tier ISCED 5A/all 5
412	General ISCED 4A/4B, access ISCED 5B/lower tier 5A
413	General ISCED 4A, access upper tier ISCED 5A/all 5
421	ISCED 4 programmes without access ISCED 5
422	Vocational ISCED 4A/4B, access ISCED 5B/lower tier 5A
423	Vocational ISCED 4A, access upper tier ISCED 5A/all 5
510	ISCED 5A short, intermediate/academic/general tertiary below
520	ISCED 5B short, advanced vocational qualifications
610	ISCED 5A medium, bachelor/equivalent from lower tier tertiary
620	ISCED 5A medium, bachelor/equivalent from upper/single tier
710	ISCED 5A long, master/equivalent from lower tier tertiary
720	ISCED 5A long, master/equivalent from upper/single tier tertiary
800	ISCED 6, doctoral degree
5555	Other

ESS: Ex-ante output harmonized

Germany

*Was ist der höchste
allgemeinbildende
Schulabschluss, den
Sie
erreicht haben?*

Grundschule nicht beendet	0
(Noch) kein Abschluss , aber Grundschule beendet	1
Abschluss einer Förderschule (Sonderschule, Hilfsschule)	2
Volks- oder Hauptschulabschluss bzw. Polytechnische Oberschule der ehem. DDR mit Abschluss der 8. oder 9. Klasse	3
Mittlere Reife , Realschulabschluss, Fachoberschulreife Mittlerer Schulabschluss bzw. Polytechnische Oberschule der ehem. DDR mit Abschluss der 10. Klasse	4
Fachhochschulreife (Abschluss einer Fachoberschule etc.)	5
Abitur, allgemeine oder fachgebundene Hochschulreife bzw. Erweiterte Oberschule der ehem. DDR mit Abschluss 12. Klasse	6
Anderer Schulabschluss, und zwar: EINTRAGEN _____	7
(Weiß nicht)	8

ESS: Ex-ante output harmonized

Germany

Was ist der höchste Ausbildungsabschluss, den Sie erreicht haben?

<u>Kein beruflicher Ausbildungsabschluss</u>	0
<u>Betriebliche Anlernzeit</u> mit Abschlusszeugnis, aber keine Lehre; Teilfacharbeiterabschluss	1
Abschlusszeugnis nach Berufsprüfungsjahr (BGJ), Berufsfachschule (berufliche Grundkenntnisse), oder <u>1-jähriger Ausbildung an einer Schule des Gesundheitswesens</u> (medizinische Hilfsberufe)	2
Abschlusszeugnis nach <u>2- bis 3-jähriger Ausbildung an einer Schule des Gesundheitswesens</u> (medizinische Assistenten, Krankenschwestern/-pfleger)	3
Berufsqualifizierender Abschluss einer <u>Berufsfachschule/eines Kollegs</u> (schulische Berufsausbildung)	4
Abschluss einer Ausbildung <u>zum Erzieher/zur Erzieherin</u>	5
Abgeschlossene <u>gewerbliche Lehre</u> /duale Ausbildung in Industrie, Handwerk oder Landwirtschaft (Facharbeiter- oder Gesellenbrief)	6
Abgeschlossene <u>kaufmännische Lehre</u> /duale Ausbildung (Kaufmannsgehilfenbrief, IHK-Prüfungszeugnis)	7
Laufbahnprüfung für den <u>mittleren Dienst</u>	8
Abschluss einer <u>2. Berufsausbildung</u> (berufliche Zweitausbildung)	9
<u>Meister-/Techniker-</u> oder gleichwertiger <u>Fachschulabschluss</u> (inkl. Fachschule der ehemaligen DDR); Abschluss einer Verwaltungs- und Wirtschaftsakademie (<u>VWA</u>) oder <u>Fachakademie</u> (Bayern)	10
Laufbahnprüfung für den gehobenen Dienst	11
2. Staatsexamen	12
Anderer Studienabschluss, und zwar: _____ (offen)	13
(Weiß nicht)	88

ESS: Ex-ante output harmonized

Germany

Was ist der höchste Studienabschluss, den Sie erreicht haben?

<u>Kein Hochschulabschluss</u>	0
Zwischenprüfung, Vordiplom	1
<u>Diplom</u> einer <u>Berufsakademie</u>	2
<u>Bachelor</u> einer <u>Verwaltungs-/Fachhochschule, Berufsakademie</u> (auch duale Hochschule Baden-Württemberg)	3
<u>Bachelor</u> einer <u>Universität</u> (auch Kunst-, Musik-, technische, theologische oder pädagogische Hochschule)	4
<u>Diplom</u> einer <u>Verwaltungs-/Fachhochschule</u> (FH, auch frühere Ingenieurschule)	5
<u>Master</u> einer <u>Verwaltungs-/Fachhochschule</u> (FH) (auch duale Hochschule Baden-Württemberg)	6
<u>Master</u> oder Aufbaustudium einer <u>Universität</u> (auch Kunst-, Musik-, technische, theologische oder pädagogische Hochschule)	7
<u>Diplom, Magister Artium</u> oder <u>1. Staatsexamen</u> einer Universität (auch Kunst-, Musik-, technische, theologische oder pädagogische Hochschule)	8
Promotion oder Habilitation	9
Anderer Studienabschluss, und zwar: _____ (offen)	10
(Weiß nicht)	88

ESS: Ex-ante output harmonized

France

Quel est le plus haut niveau d'études ou de formation que vous avez terminé ?

- 1) A - Non scolarisé ou école primaire non achevée
- 2) B - Ecole primaire uniquement
- 3) C - Certificat d'études primaires
- 4) D - Scolarité suivie de la 6ème à la 3ème
- 5) E - Brevet élémentaire, Brevet d'étude du premier cycle, Brevet des collèges
- 6) F - Scolarité suivie de la 2nde à la Terminale
- 7) G - CAP, BEP, examen de fin d'apprentissage artisanal
- 8) H - Diplôme d'aide-soignante, auxiliaire de puériculture, aide médico-pédagogique, aide à domicile
- 9) I - Baccalauréat professionnel, Brevet de technicien
- 10) J - Baccalauréat technologique, Baccalauréat de technicien, BEA, BEC, BEI, BES
- 11) K - Baccalauréat général, Brevet supérieur
- 12) L - Diplôme de la capacité en droit, Diplôme d'accès aux études universitaires (DAEU)
- 13) M - Diplôme de moniteur-éducateur, Educateur technique spécialisé, Brevet Professionnel
- 14) N - Diplôme universitaire du premier cycle (DEUG), Classes préparatoires aux grandes écoles
- 15) O - Diplôme universitaire de technologie (DUT), Brevet de technicien supérieur (BTS)
- 16) P - Certificat d'aptitude pédagogique (instituteur), Diplôme d'éducateur spécialisé, Diplôme d'assistante sociale, Diplôme paramédical (laborantin, infirmier, etc...)
- 17) Q - Licence professionnelle
- 18) R - Licence
- 19) S - Diplôme d'école d'ingénieur
- 20) T - DESS, Master deuxième année professionnel
- 21) U - Maîtrise, CAPES, CRPE (professeur des écoles)
- 22) V - DEA, DES, Master deuxième année recherche, Agrégation
- 23) W - Diplômes professionnels supérieurs divers (notaire, architecte, vétérinaire, journaliste...)
- 24) X - Diplôme des grandes écoles (hors écoles d'ingénieur)
- 25) Y - Doctorat en médecine ou équivalents (Médecine, Dentaire, Pharmacie, Vétérinaire)
- 26) Z - Doctorat
- 27) [Refus]
- 28) [Ne sait pas]

ESS: Ex-ante output harmonized

Germany

EDUBDE1=5 and EDUADE3=0 and EDUADE2=0	312 General ISCED 3A/3B, access ISCED 5B/lower tier 5A
EDUBDE1=6 and EDUADE3=0 and EDUADE2=0	313 General ISCED 3A, access upper tier ISCED 5A/all 5
EDUBDE1<5 and EDUADE3=1 and EDUADE2=0) or (EDUBDE1<5 and EDUADE3=3 and EDUADE2=0) or (EDUBDE1<5 and EDUADE3=8 and EDUADE2=0)	321 Vocational ISCED 3C >= 2 years, no access ISCED 5
(EDUBDE1<5 and EDUADE3=6 and EDUADE2=0) or (EDUBDE1<5 and EDUADE3=7 and EDUADE2=0) or (EDUBDE1<5 and EDUADE3=4 and EDUADE2=0)	322 Vocational ISCED 3A/3B, access 5B/lower tier 5A
	323 Vocational ISCED 3A, access upper tier ISCED 5A/all 5
	412 General ISCED 4A/4B, access ISCED 5B/lower tier 5A
	413 General ISCED 4A, access upper tier ISCED 5A/all 5
	421 ISCED 4 programmes without access ISCED 5
(EDUBDE1<5 and EDUADE3=9 and EDUADE2=0) or (EDUBDE1=5 and range(EDUADE3,1,8) and (EDUADE2=0 or EDUADE2>10))	422 Vocational ISCED 4A/4B, access ISCED 5B/lower tier 5A

ESS: Ex-ante output harmonized

France

F15 (EDLVDFR) = 10, 11	313 General ISCED 3A, access upper tier ISCED 5A/all 5
F15 (EDLVDFR) = 07, 08	321 Vocational ISCED 3C >= 2 years, no access ISCED 5
	322 Vocational ISCED 3A/3B, access 5B/lower tier 5A
F15 (EDLVDFR) = 09	323 Vocational ISCED 3A, access upper tier ISCED 5A/all 5
	412 General ISCED 4A/4B, access ISCED 5B/lower tier 5A
F15 (EDLVDFR) = 12	413 General ISCED 4A, access upper tier ISCED 5A/all 5
F15 (EDLVDFR) = 13	421 ISCED 4 programmes without access ISCED 5
	422 Vocational ISCED 4A/4B, access ISCED 5B/lower tier 5A

For more detailed coding and harmonization explanations on the education variable in ESS see „ESS8 Appendix A1 Education ed 2.2“, online:

<https://www.europeansocialsurvey.org/data/download.html?r=8>

ISSP: Harmonization process

The Methodology Committee co-ordinates the work of six groups addressing different areas of **cross-cultural methods**, all concerned with issues of equivalence:

- demography
- non-response
- weighting
- mode effects
- questionnaire design
- translation

ISSP: Harmonization process

- Common source questionnaire, occasionally leaving space for (country-/ culture-specific) interpretation
- Since 2001: ex-ante output for background variables, often using **international standard classifications**, such as the ISCED-97 for education
 - Followed by expertise training, careful monitoring and documentation process

ISSSP: Input harmonized

Source:

Q. 6. How much do you agree or disagree with the following statements? (Please, check one box on each line)

	Agree strongly	Agree	Neither agree nor disagree	Disagree	Disagree strongly	Can't choose
a. Large international companies are doing more and more damage to local businesses in [COUNTRY].	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

France:

6. Etes-vous d'accord ou pas d'accord avec les propositions suivantes ?

(Entourez un seul chiffre par ligne)

	Tout à fait d'accord	Plutôt d'accord	Ni d'accord, ni pas d'accord	Plutôt pas d'accord	Pas du tout d'accord	Ne peut choisir
Les grands groupes internationaux font de plus en plus de tort aux entreprises locales en France	1	2	3	4	5	8

UK:

Q6 How much do you agree or disagree with the following statements?

PLEASE TICK **ONE** BOX
ON EACH LINE

	Agree strongly	Agree	Neither agree nor disagree	Disagree	Disagree strongly	Can't choose
a. Large international companies are doing more and more damage to local businesses in Britain	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Spain:

P.48. ¿Hasta qué punto está Ud. de acuerdo o en desacuerdo con las siguientes afirmaciones?

ENTREVISTADOR: MOSTRAR TARJETA

	Muy de acuerd o	De acuerd o	Ni de acuerdo ni en desacuerdo	En desacuerd o	Muy en desacuerd o	NS	NC
a. Las empresas multinacionales están haciendo cada vez más daño a las empresas nacionales en España.	5	4	3	2	1	8	9

ISSSP: Ex-ante harmonized

Spain:

	BVQ_04 Entrevistado	BVQ_04a Cabeza de Familia
- No sabe leer ni escribir	0	0
- Sin estudios (Estudios primarios sin terminar)	1	1
- Primer grado (Certificado escolar, EGB 1ª etapa, más o menos 10 años)	2	2
- Segundo grado 1er ciclo (Graduado escolar, EGB 2ª etapa, 1º y 2º ESO, hasta los 14 años)	3	3
- Segundo grado 2º ciclo (FP 1º y 2º, Bachiller superior, BUP, 3º y 4º ESO, COU, PREU, 1º y 2º Bachillerato, hasta los 18 años)	4	4
- Tercer grado 1er ciclo (Ingeniero técnico, escuelas Universitarias, arquitectos técnicos, peritos, magisterio, ATS, diplomados, graduados sociales, asistentes sociales, etc.)	5	5
- Tercer grado 2º ciclo (Universitarios, licenciado superior, facultades, escuelas técnicas superiores, doctorados, etc.)	6	6
- Otros (especificar)	7	7
- No sabe	98	98
- No contesta	99	99

France:

35. Quel niveau d'études le plus élevé votre conjoint ou partenaire a-t-il/elle atteint ?

- Aucun	1
- Ecole primaire <u>sans</u> le certificat d'études	2
- Ecole primaire <u>avec</u> le certificat d'études	3
- Collège (de la 6^{ème} à la 3^{ème})	4
- Enseignement professionnel après le collège <u>sans</u> CAP, BEP ni baccalauréat	5
- Enseignement professionnel après le collège, <u>avec</u> CAP ou BEP, <u>sans</u> bac	6
- Enseignement professionnel après le collège <u>avec</u> baccalauréat	7
- Enseignement général des lycées (de la seconde à la terminale) <u>sans</u> baccalauréat ...	8
- Enseignement général des lycées (de la seconde à la terminale) <u>avec</u> baccalauréat....	9
- Premier cycle universitaire (ou équivalent à Bac+3)	10
- Deuxième ou troisième cycle universitaire (équivalent à Bac+4 et au-delà)	11

- Afterwards harmonized according to ISCED-97
- http://www.unesco.org/education/information/nfsunesco/doc/isced_1997.htm

EVS: Ex-ante harmonized

Religious denomination: Target Variable

Response Options and coding:

1:	Roman catholic
2:	Protestant
3:	Free church/Non-conformist/Evangelical
4:	Jew
5:	Muslim
6:	Hindu
7:	Buddhist
8:	Orthodox
9:	Other
55:	design specific missing values
77:	not applicable
88:	don't know (spontaneous)
99:	no answer (spontaneous)

EVS: Ex-ante harmonized

Religious denomination: Harmonization Mapping

Denmark

Target: v52	Source: v52_DK
2 Protestant	1 People's Church
5 Muslim	2 Muslim
7 Buddhist	3 Buddhist
6 Hindu	4 Hindu
4 Jew	5 Mosaic
9 Other	6 Asatro
9 Other	7 Other
77 not applicable	77 not applicable
88 don't know (spontaneous)	88 don't know (spontaneous)
99 no answer (spontaneous)	99 no answer (spontaneous)

Serbia

Target: v52	Source: v52_RS
8 Orthodox	1 Serbian Orthodox
1 Roman catholic	2 Roman Catholic
5 Muslim	3 Islam
4 Jew	4 Jews
2 Protestant	5 Protestant
2 Protestant	6 Evangelic
2 Protestant	7 Slovakian evangelic
2 Protestant	8 Adventist
9 Other	9 Jehovah's witnesses
2 Protestant	10 Baptist
9 Other	30 Other, please specify (Write in)
77 not applicable	77 not applicable
88 don't know (spontaneous)	88 don't know (spontaneous)
99 no answer (spontaneous)	99 no answer (spontaneous)

<https://www.gesis.org/angebot/daten-analysieren/internationale-umfragen/european-values-study/5th-wave-2017>

Examples of large harmonization projects

- The IPUMS Health Surveys: National Health Interview Survey merged and harmonized data from 1963 to present <https://nhis.ipums.org/nhis/index.shtml>
- The Eurobarometer Survey Series collects data since 1970 harmonizing it into single cross-national files <https://www.gesis.org/eurobarometer-data-service/survey-series/>
- The Multinational Time Use Study (MTUS) brings together data from 55 years and over 30 countries <https://www.timeuse.org/mtus>

International Standards

- Basis for harmonizing (international) data
- Compromise on measurements on cross-national and –cultural level
- Various examples often accompanied by discussions and modifications on concept
- Two Examples:
 - Occupational Status (ISCO)
 - Education (Hoffmeyer-Zlotnik/Warner matrix)



International Standards

- **Pro:** unambiguous concept, including coding and classification instructions → facilitates collection of comparative data
- **Con:** intense training of staff, evaluation and improvement of automated procedures and quality checks → high costs



ISCO = International Standard Classification of Occupations

- Prime example of standard classification that applies in input harmonized surveys
(Hoffmeyer-Zlotnik and Warner, 2011)
- Proposed by International Labor Organisation (ILO)
- Providing version updates: ISCO-58, -68, -88, -08
- Four-level hierarchically structured classification: classify jobs around the world in 436 units which can be classified into 130 minor, 43 sub-major groups and 10 major groups

ISCO = International Standard Classification of Occupations

Example: Police Officers = 5412

- 1st digit, major group: **5**, "Services and Sales Workers"
- 2nd digit, sub-major group: **54**, "Protective Services Workers"
- 3rd digit, minor group: **541**, "Protective Services Workers"
- 4th digit, unit: **5412**, "Police Officers"

ISCO = International Standard Classification of Occupations

- For all codes see “ISCO-08 Part 2: Classification Structure”
- <http://www.ilo.org/public/english/bureau/stat/isco/index.htm>

Other Occupation/Socioeconomic status Standards

- Duncan Socioeconomic Index (SEI)

Duncan, O. D. (1961). A Socioeconomic Index for all Occupations. In J. Reiss, Jr. (Ed.), *Occupations and Social Status* (pp. 109–138). New York: Free Press of Glencoe.

- Standard International Occupational Prestige Scale (SIOPS)

Treiman, D. J. (1977), *Occupational Prestige in Comparative Perspective*, Academic Press, New York.

- EGP class categories

- Erikson, R., J. H. Goldthorpe & L. Portocarero, (1979). Intergenerational class mobility in three Western European societies: England, France and Sweden. *The British Journal of Sociology*, 30(4), 415-441.

Measuring Occupation

- For more details about occupation standards:

Hoffmeyer-Zlotnik, J. H., & Warner, U. (2011). *Measuring Occupation and Labour Status in Cross-National Comparative Surveys*. Bonn: GESIS Leibniz Institute for the Social Sciences.



Hoffmeyer-Zlotnik/Warner matrix of Education

- Two dimensions: general education and professional education
- Requires a two-step questionnaire asking about both dimensions
- Both dimensions are brought together into one *prestige score*
 - “bring the combination from general and vocational degree together with the social prestige that a person can gain on the labour market” (p. 6)
- For coding scheme, examples of application and comparison to other indexes see p. 7 ff.

Hoffmeyer-Zlotnik/Warner matrix of Education

Table 11: Hoffmeyer-Zlotnik/Warner matrix of education – Principles.

vocational education	ISCO major groups	general education – grades, no certificates				
		non	basic degree	second degree	third degree	university-entrance diploma
non	9,8	1	2	3	6	7
dual system	8,7	4	4	5	5	5
vocational school	4,5	4	4	5	5	5
vocational college	3,4	0	5	5	8	8
college of higher education	2,3	0	0	9	9	9
university	2	0	0	0	10	10

This and following table from: Hoffmeyer-Zlotnik, J. H., & Warner, U. (2007). How to Survey Education for Cross-National Comparisons: The Hoffmeyer-Zlotnik/Warner-Matrix of Education. In *Metodološki zvezki*, 4(2): 117-148.

Hoffmeyer-Zlotnik/Warner matrix of Education

Table 12: Hoffmeyer-Zlotnik/Warner matrix of education – for **Germany**.

vocational education	general education				university-entrance diploma
	non	basic degree	second degree	third degree	
Non dual system	1	2	3	6	7
vocational school	4	4	5	5	5
vocational college	4	4	5	5	5
college of higher education	0	5	5	8	8
university	0	0	9	9	9
	0	0	0	10	10

Table 13: Hoffmeyer-Zlotnik/Warner matrix of education – for **Luxembourg**.

vocational education	general education			university-entrance diploma
	non	basic degree	second degree	
non dual system	1	2	3	7
vocational school	4	4	5	5
vocational college	4	4	5	5
college of higher education	0	5	5	8
university	0	0	9	9
	0	0	0	10

Other Education Standards



- International Standard Classification of Education
United Nations, online:
<http://uis.unesco.org/en/topic/international-standard-classification-education-isced>
- Education Years
E.g. in ESS, Variable eduyrs: Years of full-time education completed

Discussions

Harmonization decisions are accompanied by various discussions often concerning...

- The choice of international standards
 - E.g. Mutchler, J. E. and D. L. Poston, JR. (1983)
- The concept of international standards
 - Hauser, R. M. and J.R. Warren (1996)
- The choice of the harmonization strategy
 - E.g. Hoffmeyer-Zlotnik, J. and U. Warner (2006)



DOCUMENTATION



Transparency in research

- “Transparency is the cornerstone of social science. Academic discourse rests on the obligation of scholars to reveal to their colleagues the data, theory, and methodology on which their conclusions rest. Unless other scholars can examine evidence, parse the analysis, and understand the processes by which evidence and theories were chosen, why should they trust—and thus expend the time and effort to scrutinize, critique, debate, or extend—existing research?” (Moravcsik, 2007)
- Given the importance of transparency to precise replication and comparable statistical results, the lack of scientific standards for documenting harmonization can be a serious obstacle.

Moravcsik, A. (2007). Transparency: The Revolution in Qualitative Research. *PS: Political Science & Politics*, 47 (1): 48–53. <http://dx.doi.org/10.1017/S1049096513001789>

Freese, J. (2007). Replication standards for quantitative social science: Why not sociology. *Sociological Methods Research*, 36: 153–172. <https://journals.sagepub.com/doi/pdf/10.1177/0049124107306659>

Why care about documentation?

Crucial for...

- **Evaluating** data (improvement of *ex-ante* harmonization)
- **Replicating/testing** study (improvement of *ex-ante*, right application in *ex-post*)
- **Re-using** data = secondary analysis (right application in *ex-post*)

Why care about documentation?

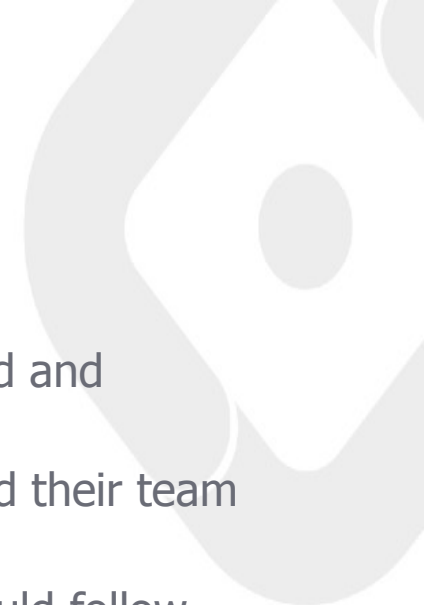
Pressure to document increasingly driven by

- more and better **quality** data available to harmonize;
- increasingly sophisticated statistical **methods** used to analyse said data;
- and specific information required to **replicate** variables, thus advancing scientific theories.

→ Not only is replication key to evaluating the quality of a piece of research, but it allows new research paths to be developed and explored

How and what?

- Ideally, **from beginning** of survey life cycle and in most detailed and comprehensible way for all steps
- Often most efforts are put on data collection, but researchers and their team should also be **aware** of the importance of documenting
- Consider documentation as integral part of study design, i.e. should follow **disclosure requirements** and not just refer to what is available



How and what?

- Publication and if necessary, later versions (1.0, 2.0) or errata of harmonized datasets and accompanying documentation
 - All data conversion made should be reversible
- Integrate and unify documentation tools
 - Avoid losing information in E-Mails or text editors
 - Documentation software can help!



Code alone is insufficient



- Including the syntax or do file with an article does not solve the problem:
 - there are no conventions that outline how the syntax code should be published (i.e. SPSS syntax or Stata do files);
 - code alone may not contain the information necessary to replicate others' work (i.e. if questions were answered based on prior routing); and
 - providing code may not include how certain coding decisions were made.

How and what?

- Given the importance of transparency to precise replication and comparable statistical results, the lack of scientific standards for documenting harmonization can be a serious obstacle
- Different authors aim to bring forward a standard framework for documentation
- Mohler and Uhler (2003) propose a framework of elements for general study documentation that can be applied to harmonization documentation as well ...



How and what?

- A comprehensive description and explanation of a **project's theoretical goals**;
- Documentation of the **operationalization** of theoretical concepts, including how **concepts** *map* to latent **constructs** and constructs *relate* to **indicators**, as well as documentation of **translations** and any **adaption** of a source questionnaire;
- A detailed description of all the relevant aspects of **design, implementation, and outcomes**, plus **assessment of the quality** of the outcomes;

How and what?

- A 'thick' description of the **socio-cultural contexts** of a cross-national study and the different 'fields', including all the relevant contextual statistics. Some of these are standard and basic (e.g., population statistics), others are determined by the scope and topic of a study (e.g., divorce rates or GNP). This information should also cover **regulations** and **infrastructure information** relevant for understanding respondent answers in a given context (e.g., divorce laws, abortion regulations)

How and what?

Specially important in harmonization:

- *Cultural* context as well as *linguistic* meanings need to be strictly respected when mapping (meta)data
- Loss of information through merging items can create bias in harmonized data



Excursus: Loss of information

- When harmonizing information can be lost through merging categories or modifying labels
- “Harmonizers” should be aware of this and sometimes must decide on a trade-off

Excursus: Loss of information

Multi-digit coding in harmonized US census without information loss:

- In 1920, asked for adopted children; in 1940, not asked for it
- Indicating relation to head of household: 03 = child
- Indicating adopted child: 02 = adopted
- 0302 = adopted child, 0301 = child (preserved information on adoption)

Excursus: Loss of information

Trade union membership in ISSP as example with information loss:

- Current member Member
- Former member Nonmember
- Never member Nonmember

How and what?

Key elements of documentation harmonization according to Granda et al. (2010):

- Target variables, with information from source and transformation algorithm
- Provide original data files, original questionnaire, information about data collection
- Harmonization syntax/code



How and what?

- Complete set of documentation including crosswalks describing relation between source variables and counterparts
- User's Guide: idiosyncrasies and special characteristics of targets and components
- Summary of data processing steps
- Full specification of recoded variables



Which metadata, information is necessary and sufficient to replicate?

Winters and Netscher (2016) determined the following metadata were necessary to report in order for someone else to have sufficient information to recreate the harmonized variable:

- Name of the project and its author(s) so the work is searchable and findable,
- Harmonization code (in at least both SPSS and Stata) for precise replication,
- Information on the purpose of the harmonization,
- The name of the concept to be operationalized and harmonized, literature or previous work that informs the recoding,

Which metadata, information is necessary and sufficient to replicate?

- **Target and source variables metadata** including name, label, definition, the response options values and their labels as well as the level of measurement and whether it is individual or aggregate level data,
- **Target and source question wording metadata**, including notes on the questions and relevant instructions to interviewers on how the question was to be administered, and
- **Study metadata** and its permanent identifier or online source for the data

How and what?



- A perfect documentation also includes:
 - Access to the original data (if possible)
 - Original questionnaire wording
 - Information about survey collection process
 - Assistance for check-backs or re-transformation plans of users
 - Restricted-use data agreements to allow access under controlled conditions (deposit in a Secure Data Center, original data might be in a SDC)

See links to FAIR principles in FAIR lecture

How and what?

- A complete data management plan checklist also including an orientation for metadata is provided by CESSDA:
https://www.cessda.eu/content/download/4302/48656/file/TTT_DMO_DMPExpertGuide_v1.2.pdf

Metadata

- Documenting basically means collecting metadata which is „**data about your data**“
- Metadata are descriptors that facilitate cataloguing data and data discovery
- Can be stored in a data repository as well and be transformed into machine-readable metadata

Metadata

- Metadata can be found everywhere – also outside research and the social sciences
- Core of most software-packages that we use in our everyday life
- Following slide shows some software examples. Think of possible metadata relevant to them!



Metadata



WIKIPEDIA
The Free Encyclopedia



Google

You Tube



Metadata

- Riley (2017) distinguishes between 6 types of metadata categorised in 4 groups

Types of Metadata

Descriptive metadata	For finding or understanding a resource
Administrative metadata <ul style="list-style-type: none">- Technical metadata- Preservation metadata- Rights metadata	<ul style="list-style-type: none">- For decoding and rendering files- Long-term management of files- Intellectual property rights attached to content
Structural metadata	Relationships of parts of resources to one another
Markup languages	Integrates metadata and flags for other structural or semantic features within content

"Types of metadata" in Riley, J. (2017). Understanding Metadata. What is Metadata, and what is it for? Baltimore: National Information Standards Organization (NISO).

Metadata

Metadata Type	Example Properties	Primary Uses
Descriptive metadata	Title Author Subject Genre Publication date	Discovery Display Interoperability
Technical metadata	File type File size Creation date/time Compression scheme	Interoperability Digital object management Preservation
Preservation metadata	Checksum Preservation event	Interoperability Digital object management Preservation
Rights metadata	Copyright status License terms Rights holder	Interoperability Digital object management
Structural metadata	Sequence Place in hierarchy	Navigation
Markup languages	Paragraph Heading List Name Date	Navigation Interoperability

"Types of metadata" in Riley, J. (2017). Understanding Metadata. What is Metadata, and what is it for? Baltimore: National Information Standards Organization (NISO).

Metadata

- Machine-readable metadata support your documentation by explaining the purpose, origin, time, location, creator(s), terms of use, and access conditions of research data
- **DDI** (Data Documentation Initiative) (DDI Alliance, 2017b) is an international standard for describing the data produced by surveys and other observational methods in the social, behavioural, economic, and health sciences
- Expressed in XML, the DDI metadata specification supports the entire research data lifecycle



DDI for social sciences

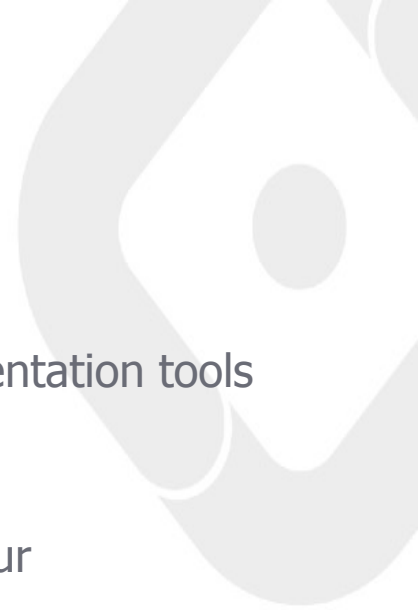
DDI (Data Documentation Initiative) (DDI Alliance, 2017b) is an international standard for describing the data produced by surveys and other observational methods in the social, behavioural, economic, and health sciences. Expressed in XML, the DDI **metadata** specification supports the entire research data lifecycle.

Common fields in the DDI include:

- Title
- Alternate Title
- Principal Investigator
- Funding
- Bibliographic Citation
- Series Information
- Summary
- Subject Terms
- Geographic Coverage
- Time Period
- Date of Collection
- Unit of Observation
- Universe
- Data Type
- Sampling
- Weights
- Mode of Collection
- Response Rates
- Extent of Processing
- Restrictions
- Version History

And now...?

- Working with documentation software and/or web-based documentation tools can really assist your work
 - E.g. CharmStats
- (Interactive) online libraries serve as a good platform to store your documentation
 - E.g. Hub



And now...?

All documentation work only becomes meaningful if you...

- Guarantee as much **transparency** as possible to creators and users of the data
- Publish and store your data and documentation in a **FAIR** way



Using Persistent Identifiers (PID)

- A PID will ensure users find and can access data and documentation (see FAIR principles)
- The identifier enables long-term preservation always referring the user to the original source
- Prevents link rot (hyperlinks to unavailable web objects)
- DOI (digital object identifier) is a well-known persistent identifier in academia
<https://www.doi.org/>

Archives

The harmonized data together with the metadata can be stored in data archives (e.g. GESIS supporting ISSP, NDS supporting ESS) providing a link to the documentation files

- Chapter 6 of CESSDA Data Management Expert Guide offers detailed information on archiving and publishing
- Including a list of CESSDA data archives

<https://www.cessda.eu/Training/Training-Resources/Library/Data-Management-Expert-Guide/6.-Archive-Publish/Publishing-with-CESSDA-archives>

Benefits of (CESSDA) Data Archives

Thanks to expert help from data archives your datasets will profit from an increased

- Comprehensibility
- Visibility
- Findability
- Reusability
- Longevity
- Overall quality

Read more: <https://www.cessda.eu/Training/Training-Resources/Library/Data-Management-Expert-Guide/6.-Archive-Publish/Publishing-with-CESSDA-archives>

Quality



Quality criteria

- Quality criteria play an important role in statistics and accordingly also in harmonization
- For Eurostat and the European Statistical System (ESS) quality has also a high standing
- ESS edited and published the *Quality Assurance Framework of the European Statistical System* on Eurostat's website to set a standard European framework
- Besides an explanation of the **principles** they provide **methods** to reach the principles

Quality criteria

ESS general statistical quality criteria comprise 12 principles divided into three areas:

Institutional environment

- Commitment to quality
- Statistical Confidentiality
- Impartiality and objectivity

Statistical processes

- Sound Methodology
- Appropriate Statistical Procedures
- Non-Excessive Burden on Respondents
- Cost effectiveness

Quality criteria

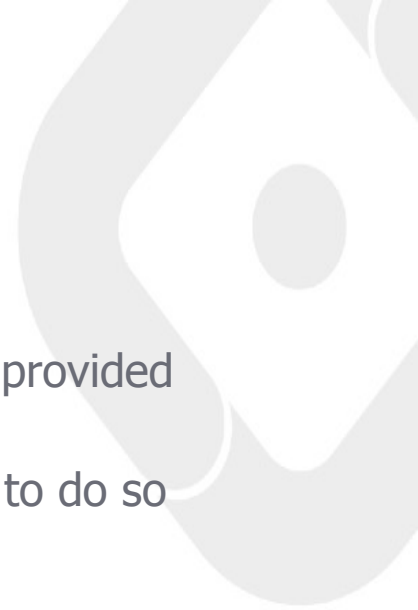
Statistical output

- Relevance
- Accuracy and Reliability
- Timeliness and Punctuality
- Coherence and Comparability
- Accessibility and Clarity

More information on quality in statistics can be found at Eurostat:
<https://ec.europa.eu/eurostat/web/quality>

Quality criteria

- Besides quality criteria for the data themselves the quality of the provided **metadata** should be screened as well
- The FAIR principles portray relatively simple but effective criteria to do so
- Again proposed standards are best to use
 - For instance, CESSDA Data Management Expert Guide including a one-paged checklist on “How FAIR are your data?”



Quality criteria

Granda and Blasczyk (2016) consider the following three as most relevant criteria in harmonization:

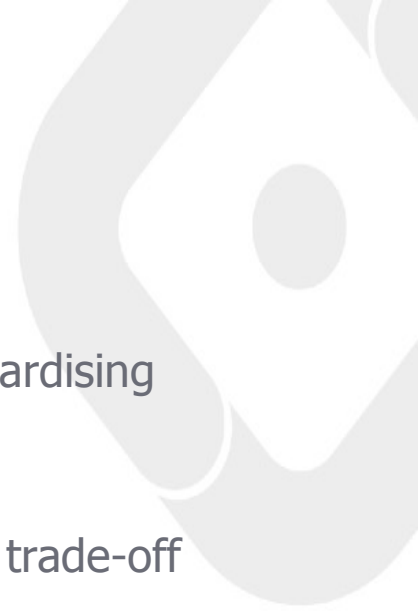
- **Consistency:** similarity of results from multiple independent harmonization efforts;
- **Completeness:** degree to which original information is preserved in harmonized data;
- **Comparability:** degree to which harmonized outputs from different statistics can report real population differences/similarities

Quality criteria

- Yet, the central criterion is **comparability** (referring to spatial and/or temporal comparability)
- In fact, the main goal and central element of the harmonization process is to achieve a higher degree of (international) comparability
- See again Granda et al. (2010): “Harmonization is a generic term for procedures aimed at **achieving or at least improving the comparability** of different surveys.”

Quality criteria

- Be aware that the highest quality is *not* achieved by simply standardising everything
 - Consider criterion of *completeness*
- There is no harmonization strategy to universally prefer, rather a trade-off regarding benefits and shortcomings needs to be done



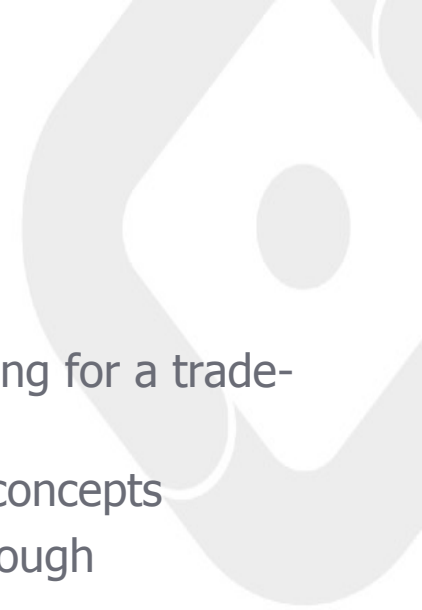
Quality criteria

- If there are too profound differences in institutional structure, it is advisable to output-harmonize the data and pay respect to the differences
- In this case *higher* quality is guaranteed by standardising *less*
- Also, a lack of flexibility due to entirely standardised surveys can lead to low acceptance by national statistical offices



Quality criteria

- It can come to a competition between different quality criteria asking for a trade-off between the different criteria
- (International) comparability vs. relevance of (national) statistical concepts
 - Domestic particularities of institutional structure can get lost through standardization
- (International) comparability vs. coherence
 - Existing national coherence on a measurement can be affected by international standardization



Quality check

- Quality checks should be involved during and after the harmonization process
- Quality criteria are an integral part and stand as requirements during the process
- That means they should be set before and kept in mind during the entire process
- It is advisable to use standard criteria such as previously seen



Quality check

- As such quality criteria not only help to judge the work but at the same time give some orientation along the process
- Again, documentation is crucial!
 - For instance, document your criteria, report on (meta)data quality, name quality conflicts



Literature

- Burkhauser**, R. V.; D. R. Lillard (2005). The Contribution and Potential of Data Harmonization for Cross-National Comparative Research. *DIW Discussion Papers*, 486. <http://hdl.handle.net/10419/18337>
- Doiron**, D., P. Raina, V. Ferretti, F. L'Heureux & I. Fortier (2012). Facilitating collaborative research: Implementing a platform supporting data harmonization and pooling. In *Norsk Epidemiologi*, 21 (2): 221-224.
- Ehling**, M. (2003). Harmonising data in official statistics. In Hoffmeyer-Zlotnik and Wolf (eds.), *Advances in cross-national comparison* (pp. 17-31). New York, Kluwer Academic/Plenum Publishers.
- Fortier**, I., D. Doiron, P. Burton & P. Raina (2011). Invited Commentary: Consolidating Data Harmonization—How to Obtain Quality and Applicability? In *American Journal of Epidemiology*, 174 (3): DOI: 10.1093/aje/kwr194
- Granda**, P., E. Blasczyk (2016). Data Harmonization. *Cross-Cultural Survey Guidelines*. Online retrievable at: <https://ccsq.isr.umich.edu/index.php/chapters/data-harmonization-chapter>
- Granda**, P., C. Wolf, R. Hadorn (2010). Harmonizing Survey Data. In Harkness et al. (eds.), *Survey Methods in Multinational, Multiregional, and Multicultural Contexts* (p. 315). New York, John Wiley & Sons, Inc.
- Harkness**, J. et al. (2003): Questionnaire design in comparative research. In Harkness, Van de Vijer, Mohler (Eds.): *Cross-cultural survey methods*, pp. 19-34. New Jersey: John Wiley & Sons, Inc.
- Harkness**, J. (2003): Questionnaire Translation. Harkness, Van de Vijer, Mohler (Eds.): *Cross-cultural survey methods*, pp. 19-34. New Jersey: John Wiley & Sons, Inc.
- Hoffmeyer-Zlotnik**, J. H. & Warner, U. (2007). How to Survey Education for Cross-National Comparisons: The Hoffmeyer-Zlotnik/Warner-Matrix of Education. In *Metodološki zvezki*, 4(2): 117-148.
- Hoffmeyer-Zlotnik**, J. H., & Warner, U. (2011). Measuring Occupation and Labour Status in Cross-National Comparative Surveys. GESIS – Leibniz Institute for the Social Sciences.
- Mohler**, P. & R. Uhler (2003): Documenting Comparative Surveys for Secondary Analysis. In Harkness, Van de Vijer, Mohler (Eds.): *Cross-cultural survey methods*, pp. 19-34. New Jersey: John Wiley & Sons, Inc.
- Riley**, J. (2017). Understanding Metadata. What is Metadata, and what is it for? Baltimore: National Information Standards Organization (NISO).
- Winters**, K., S. Netscher (2016). Proposed Standards for Variable Harmonization Documentation and Referencing: A Case Study Using QuickCharmStats 1.1. *PLOS ONE*, 11(2): e0147795. <https://doi.org/10.1371/journal.pone.0147795>