Final Conference

Adding biomedical data to the SHARE repository

Fabio Franzese, SHARE

SSHOC Final Conference
6 April 2022
Brussels/Online

social sciences & humanities open cloud

SHARE
SURVEY OF HEALTH, AGEING AND RETIREMENT IN EUROPE

- Cross-national
  - 27 European countries & Israel
- Comparable over time
- Multidisciplinary

SPONSORED BY THE

Federal Ministry of Education and Research

NIH National Institute on Aging

www.share-eric.eu

SSHOC
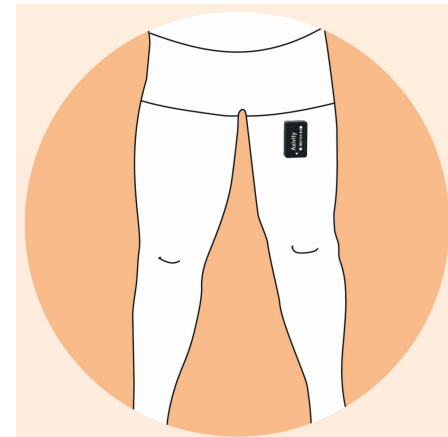social sciences & humanities open cloud
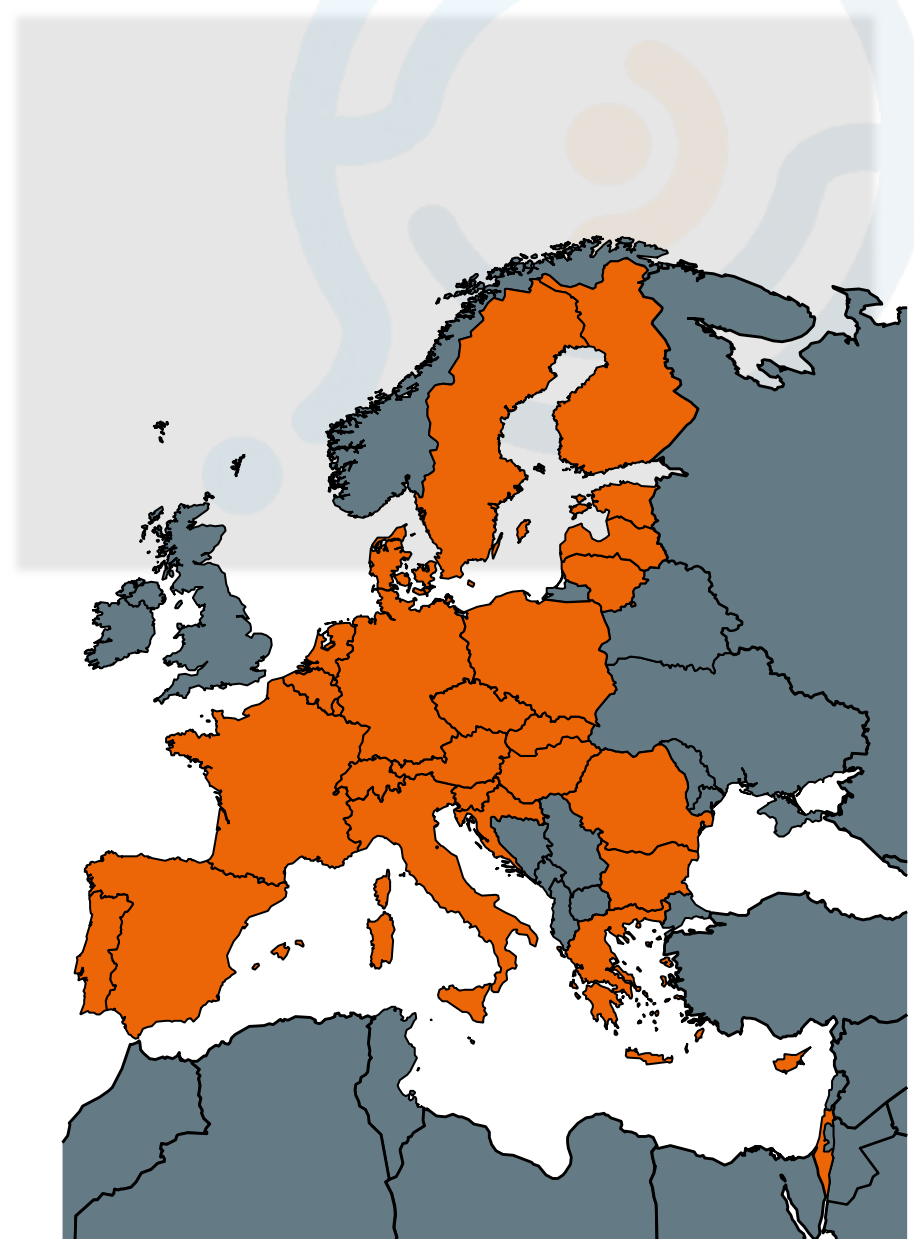
Biomedical data in SHARE

- Dried blood spots (Wave 6)
  - Drops of capillary whole blood from a simple finger prick
  - Collected on filter paper
  - Blood marker, e.g. HbA1c

© SHARE Biobank

- Accelerometry (Wave 8)
  - Measurement of physical activity
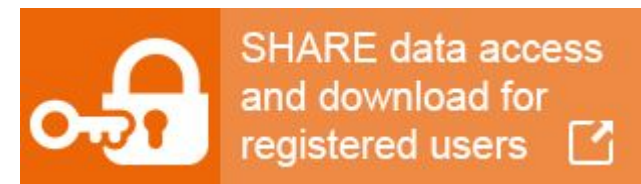 - Thigh-worn accelerometer
  - Eight days

© SHARE ERIC

SSHOC

social sciences & humanities open cloud

Repository
SHARE Research Data Center

CENTERDATA (Tilburg University, The Netherlands) provides

- Archive

- Dissemination platform

- Easy and secure access for SHARE users

SHARE data access and download for registered users

SSHOC
social sciences & humanities open cloud

Challenges & Achievements
Guidelines for ethics considerations

- 'do not harm' principle
  - include physical harm

- Review (and adapt) protection of anonymity and confidentiality

- Review by an ethics committee
  - recommended and legally required in some countries

More information: SSHOC D5.1

SSHOC
social sciences & humanities open cloud

Challenges & Achievements
FAIR data

- Findability:
  - DOI

- Accessibility:
  - Data is free of charge for scientific purpose

- Interoperability:
  - commonly-used file formats
  - (meta)data is comprehensively documented, based on DDI standards

- Reusability
  - Archive of previous release versions

More information: SSHOC D5.2 & D5.3

SSHOC
social sciences & humanities open cloud

Challenges & Achievements
Integration in SHARE repository

- First accelerometry data published in June 2021
  - Updated in February 2022

- More updates in the future
  - New metrics for accelerometry
  - Dried blood spots

More information: SSHOC D5.4

SSHOC
social sciences & humanities open cloud

Adding biomedical data to the SHARE repository

- Guidelines for ethics considerations
  - 'do not harm'
- FAIR data
  - Data access protocol
- Integration in SHARE repository
  - First data already published
  - Updates in the future

SSHOC
social sciences & humanities open cloud

Thank you for your attention!

🌐 **https://www.sshopencloud.eu**

🐦 **@SSHOpenCloud**

SHARE **https://www.share-eric.eu**

🐦 **@SHARE_MEA**

SSHOC
social sciences & humanities open cloud

Final Conference

SurveyCodings: Classifying Responses in Real Time

Rodrigo Reyes, Junior Researcher,
Tilburg University
SSHOC Final Conference
06/04/2022
Brussels

SSHOC
social sciences & humanities open cloud

www.surveycodings.org
@SurveyCodings

What is SurveyCodings?

An **online tool** for the coding/measurement of **crucial social science** variables:

- Fields and levels of education

- Occupation

- Industry

- Job tasks

- Religious Denominations

- Region

- Cost of living

Available in **different languages** for **multiple countries**

SSHOC
social sciences & humanities open cloud

What is SurveyCodings?

Jointly **carried out by**

- EVS – TiU: Ruud Luijkx, Daniela Negoita, and Rodrigo Reyes
- GESIS: Verena Ortmans and Silke Schneider
- Centerdata: Maurice Martens
- WageIndicator: Kea Tijdens

For public opinion **surveys,** social science **projects,** data **archives, researchers,** and any other interested parties

SSHOC
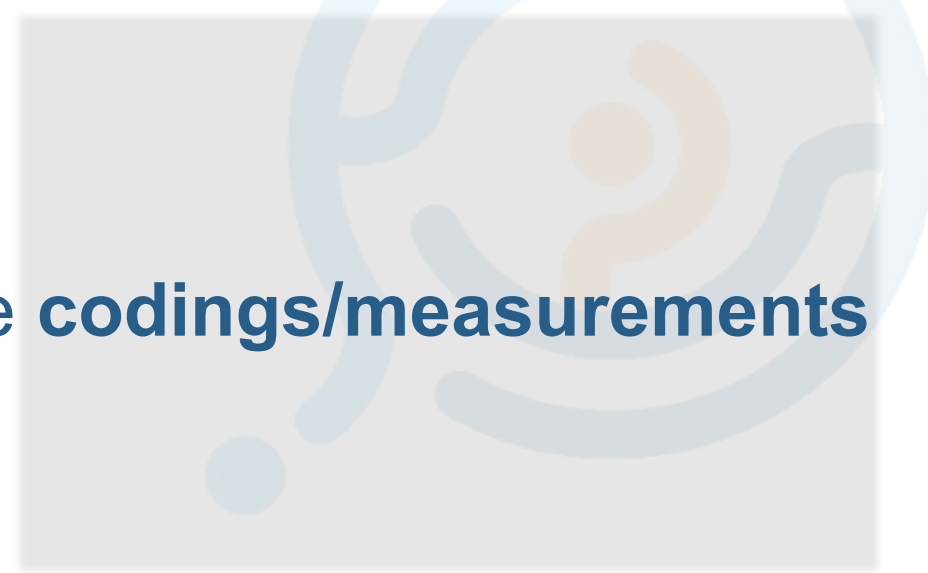social sciences & humanities open cloud

What is SurveyCodings' aim?

To **openly** offer classic socio-economic variable **codings/measurements** that are:

- **F**indable
- **A**ccessible
- **I**nteroperable
- **R**eusable

All **available** to interviewers and participants **in real time**.

SSHOC
social sciences & humanities open cloud

What does SurveyCodings offer?

	Countries	Languages	Entries	Standard
Fields of education	-	35	80	ISCED-F 2013
Levels of education	100	-	3500	ISCED 1997/2011, edulvlb
Occupation	160	54	4200	ISCO08
Industry	163	55	321	NACE coded with ISIC mapping tables
Religious denomination	211	80	497	-
Region	200	-	2000	NUTS
Cost of living	177	58	209	-

SSHOC
social sciences & humanities open cloud

What does SurveyCodings offer?

Please select a country/language

United Kingdom - England, Wales, and Northern Ireland

Search box

gradua

City and Guilds Level 6, **Gradua**teship - GCGI (NVQ level 5)
City and Guilds Level 7, Post**gradua**te (NVQ level 5)
Post-**Gradua**te Certificates
Post-**Gradua**te Diplomas

Search tree

▶ No educational qualification
Key or Basic Skills, Entry Level Certificate, Skills for Life
▶ NVQ level 1, 1-4 GCSEs A*-C, GCSEs D-G, or equivalent qualification
▶ NVQ level 2, 5 or more GCSEs A*-C, or equivalent qualification
▶ NVQ level 3 or equivalent vocational qualification
▶ A-level or equivalent general qualification
▶ NVQ level 4 and 5 or equivalent vocational qualification
▼ University degrees
    Foundation degree
    Bachelor's degree, 3-4 years
    Bachelor's degree, 5 years
    Post-Graduate Diploma or Certificate
    Master's degree
    Doctorate, PhD

Information

Description: Post-Graduate Diploma or Certificate
Classifications:

- ISCED1997: 5AI - Long ISCED 5 with access to 6
- ISCED1997-LFS: 52 - ISCED 5a
- Alternative ISCED1997: 5AI - Long ISCED 5 with access to 6
- ISCED-A2011: 760 - Master's level education, orientation unspecified
- ISCED2011-LFS: 700 - ISCED 7
- Alternative ISCED2011: 750 - Professional master's level education
- edulvlb: 710 - Vocational/professional ISCED 7

Synonyms:

What does SurveyCodings offer?

Please select a country/language

United Kingdom - England, Wales, and Northern Ireland

Search box

gradua

City and Guilds Level 6, **Gradua**teship - GCGI (NVQ level 5)
City and Guilds Level 7, Post**gradua**te (NVQ level 5)
Post-**Gradua**te Certificates
Post-**Gradua**te Diplomas

Search tree

▶ No educational qualification
  Key or Basic Skills, Entry Level Certificate, Skills for Life
▶ NVQ level 1, 1-4 GCSEs A*-C, GCSEs D-G, or equivalent qualification
▶ NVQ level 2, 5 or more GCSEs A*-C, or equivalent qualification
▶ NVQ level 3 or equivalent vocational qualification
▶ A-level or equivalent general qualification
▶ NVQ level 4 and 5 or equivalent vocational qualification
▼ University degrees
    Foundation degree
    Bachelor's degree, 3-4 years
    Bachelor's degree, 5 years
    Post-Graduate Diploma or Certificate
    Master's degree
    Doctorate, PhD

Information

Description: Post-Graduate Diploma or Certificate
Classifications:
- ISCED1997: 5AI - Long ISCED 5 with access to 6
- ISCED1997-LFS: 52 - ISCED 5a
- Alternative ISCED1997: 5AI - Long ISCED 5 with access to 6
- ISCED-A2011: 760 - Master's level education, orientation unspecified
- ISCED2011-LFS: 700 - ISCED 7
- Alternative ISCED2011: 750 - Professional master's level education
- edulvlb: 710 - Vocational/professional ISCED 7

Synonyms:

What does SurveyCodings offer?

What does SurveyCodings offer?

Please select a country/language

United Kingdom - England, Wales, and Northern Ireland

Search box

gradua

City and Guilds Level 6, **Gradua**teship - GCGI (NVQ level 5)
City and Guilds Level 7, Post**gradua**te (NVQ level 5)
Post-**Gradua**te Certificates
Post-**Gradua**te Diplomas

Search tree

▶ No educational qualification
   Key or Basic Skills, Entry Level Certificate, Skills for Life
▶ NVQ level 1, 1-4 GCSEs A*-C, GCSEs D-G, or equivalent qualification
▶ NVQ level 2, 5 or more GCSEs A*-C, or equivalent qualification
▶ NVQ level 3 or equivalent vocational qualification
▶ A-level or equivalent general qualification
▶ NVQ level 4 and 5 or equivalent vocational qualification
▼ University degrees
    Foundation degree
    Bachelor's degree, 3-4 years
    Bachelor's degree, 5 years
    Post-Graduate Diploma or Certificate

Information

**Description**: Post-Graduate Diploma or Certificate
Classifications:
- ISCED1997: 5AI - Long ISCED 5 with access to 6
- ISCED1997-LFS: 52 - ISCED 5a
- Alternative ISCED1997: 5AI - Long ISCED 5 with access to 6
- ISCED-A2011: 760 - Master's level education, orientation unspecified
- ISCED2011-LFS: 700 - ISCED 7
- Alternative ISCED2011: 750 - Professional master's level education
- edulvlb: 710 - Vocational/professional ISCED 7

Synonyms:

Synonyms:

The future of SurveyCodings

- To be **continued after SSHOC**
- Constant maintenance to keep our variables up to date
  - Standard classifications change
- Non-European languages curation
- Covering more languages

SSHOC
social sciences & humanities open cloud
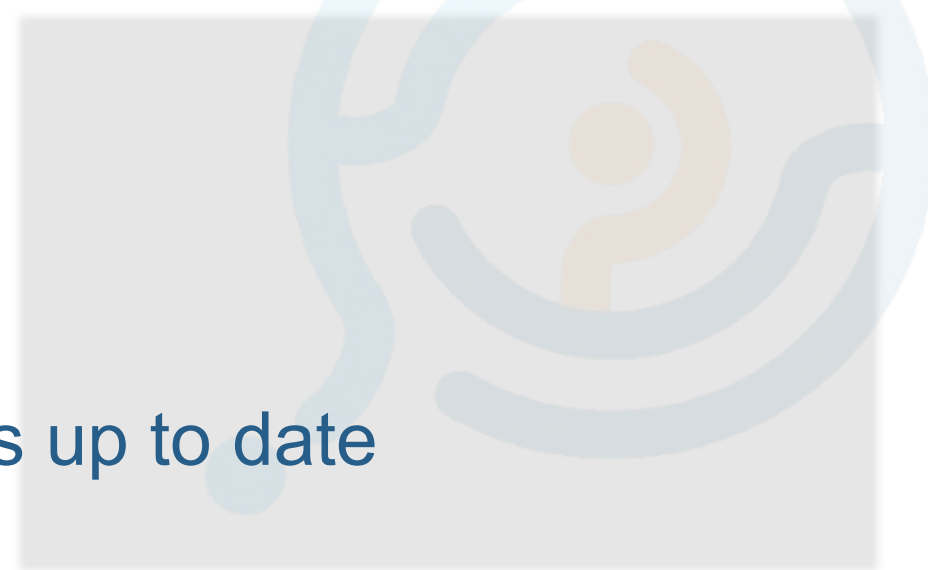
Why SurveyCodings?

- Input and ex-ante output **harmonization**
  - **Reduced workload** for surveys and post-coders
  - Enables cross-national data collection and **comparability**
  - Enables **inter-survey** cooperation
  - Updated **standard classifications**
- A **field-deployable live-search tool** for **each variable**
  - Looking at single country codings
  - Linking coding sets
  - Our website's functionality can be **stored as an app** on mobile devices for **F2F interviews**

SSHOC
social sciences & humanities open cloud
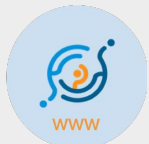
Thank you for your attention!

SurveyCodings

www.surveycodings.org

Join our community

https://www.sshopencloud.eu

@SSHOpenCloud

info@sshopencloud.eu

/in/sshopencloud

SSHOC
social sciences & humanities open cloud

[MCSQ] The Multilingual Corpus of Survey Questionnaires

European Social Survey

upf. Universitat Pompeu Fabra Barcelona

RECSM Research and Expertise Centre for Survey Methodology

MØREFORSKING

Lidun Hareide, Møreforsking Research Institute, Norway

SSHOC final Conference

April 06, 2023

SSHOC

social sciences & humanities open cloud

What is a corpus?

- A collection of
  1. ***Machine readable***
  2. ***Authentic* texts**
  3. ***Sampled* to be**
  4. ***Representative* of a particular language/language variety/domain (e.g. *Literary works, medical texts, survey questionnaires etc.*)**

- «A significant advantage of the corpus linguistic method is that it allows for the analyst to approach the study of language from the context for the scientific method» - Geoffrey Leech

SSHOC
social sciences & humanities open cloud

MCSQ: the first publicly available corpus of survey questionnaires

- **Version 3 (Rosalind Franklin):**
- **306 distinct questionnaires from:**
  - the European Social Survey (ESS),
  - the European Values Study (EVS),
  - Survey of Health Ageing and Retirement in Europe (SHARE)
  - WageIndicator Survey

  – More than 4 million words
  – ≅ 766.000 sentences

The questionnaires included comprise more than 40 years of survey research from large-scale comparative survey projects that provide cross-national and cross-cultural data to the Social Sciences and Humanities (SSH)

Qualities of the [MCSQ]:

- Open access
- Searchable
- Sentence aligned
- annotated (Pos-tagging and Name Entity Recognition)
- **FAIR** (Findable, Accessible, Reproducible and Interoperable) by design

- A powerful instrument for the further development of **best practice in design of source questionnaire and questionnaire translation methodologies**

SSHOC

social sciences & humanities open cloud

Languages included in the MCSQ

Source language: **English localized for Great Britain**

- **8 target languages adding up to 30 language varieties**:

  - Catalan
  - Czech
  - French   localized for France, Switzerland, Belgium and Luxembourg
  - German   localized for Austrian, German, Swiss and Luxembourg
  - Norwegian   localized for Bokmål
  - Portuguese   localized for Portugal and Luxembourg
  - Spanish   localized for Spain
  - Russian   localized for Azerbaijan, Belarus, Estonia, Georgia, Israel, Latvia, Lithuania, Moldavia, Russia and Ukraine

- The MCSQ is representative of the **specialized language of surveys** in the 8 languages, but not of the 8 languages in general

SSHOC

social sciences & humanities open cloud

Why is the MCSQ needed?

Ask the same question (**ASQ method**) is at the heart of large-scale survey projects that provide cross-national and cross-cultural data for the Social Sciences and Humanities

- Any translation is expected to produce texts that are *functionally equivalent* for the purpose of statistical analysis.

- Concepts to be measured must be kept *the same* across languages:
  - Keep the same psychometric properties and capture the same psychological variables (e.g. opinions and attitudes) across linguistic contexts (Harkness et al., 2010; Mohler & Johnson, 2010, Zavala-Rojas et al., 2018)
  - Translation teams should minimize adaptation.
  - Low quality translations hamper data comparability and increase errors of measurement (Davidov & De Beuckelaer, 2010; Oberski et al., 2007).

SSHOC
social sciences & humanities open cloud
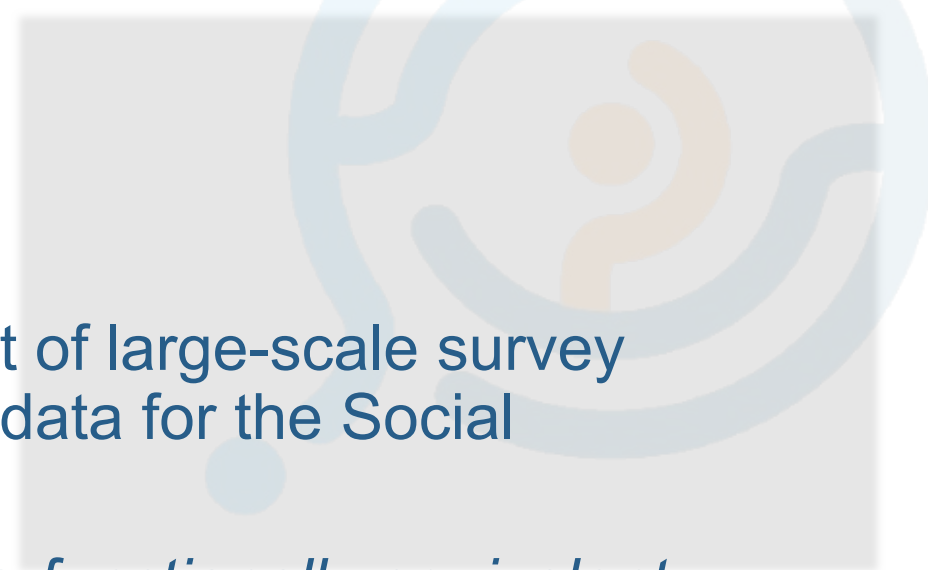
TRAPD method: the gold standard in survey translation

For translation in the ASQ framework, the Translation, Review, Adjudication, Pretesting and Documentation method (TRAPD, Harkness 2003) is used.

- Questionnaires in the MCSQ were translated using the TRAPD method.

Challenges:

- Human work intensive

- Translations are not necessarily harmonized across languages
  - Variations may reflect the teams choices and not necessarily linguistic differences
    - May hamper data comparability
    - Translation options multiply – hindering replicability
    - Managing, storing, analysing and reusing translation documentation is challenging

SSHOC

social sciences & humanities open cloud
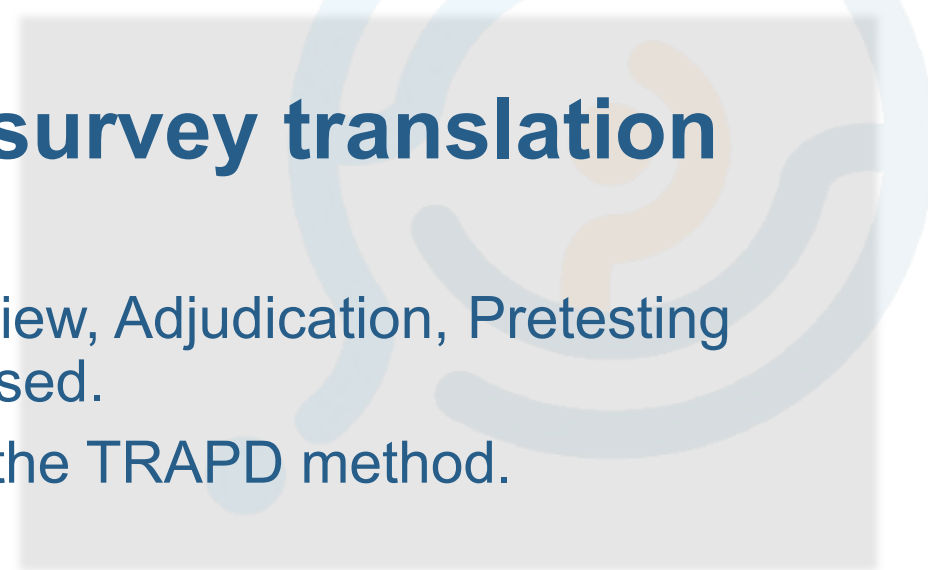
How can the MCSQ contribute to survey translation?

- **Searchable database**
  - Facilitates visualization and statistical analysis of previous translation decisions across languages
  - Tool for checking the translation of concepts across languages and surveys
- **Repository for previous rounds/waves of surveys**
  - Allows for the retrieval and preservation of source and translated questionnaires
  - Provides textual data for survey translation activities and research
  - Allows for the integration of translation analysis into the design of the source questionnaire
- **Valuable database for training new survey designers and translators and for research**
- **Can be downloaded as a translation memory and used in a Computer Assisted Translation (CAT) tool**

SSHOC

social sciences & humanities open cloud

Example of inconsistencies in translation that may hamper data comparability (from the MCSQ)

Example:

- **Most people can be trusted.** (ESS R06)

- **(BE)** *La plupart des personnes sont dignes de confiance. [Lit] (Most people are trustworthy.)*

- **(CH)** *On peut faire confiance à la plupart des personnes. [Lit] (One can trust most people.)*

- **(FR)** *On peut faire confiance aux gens. [Lit] (One can trust people.)*

A more standardized approach to translation across countries and languages is needed to enhance comparability.

- The MCSQ was created to contribute to this end

SSHOC
social sciences & humanities open cloud
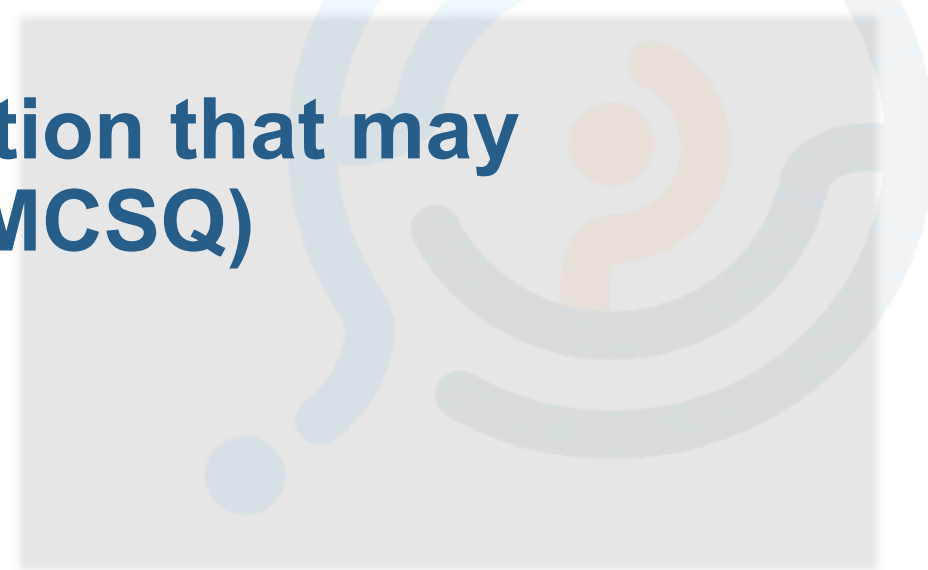
# MCSQ is also a valuable resource for

- Regional languages and language varieties (i.e. Catalan, Norwegian Bokmål, Swiss German)
- Facilitates visualization and statistical analysis of previous translation decisions across languages

- Cross-linguistic comparison of specialized use of survey language

- The creation of translation memories (TMX format)
  - Can be built and downloaded directly from the interface
  - Compatible with Computer Assisted Translation tools (Matecat)

SSHOC
social sciences & humanities open cloud

The MCSQ's successful dissemination program

- 9 presentations in international conferences
- 2 academic papers
- 1 science communication blog

SSHOC

social sciences & humanities open cloud
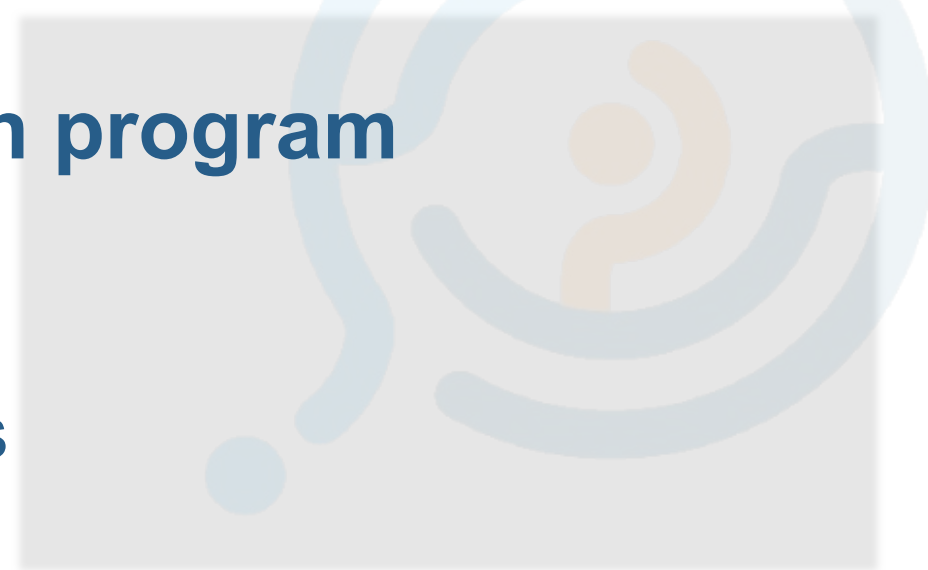
Where do we go from here?

- A corpus can be:
  - a snapshot a linguistic reality (synchronic)
  - a resource that grows more valuable over time (diachronic)

- Today the MCSQ is a really well-made snapshot of
- 306 distinct questionnaires from:
  - the European Social Survey (ESS),
  - the European Values Study (EVS),
  - Survey of Health Ageing and Retirement in Europe (SHARE)
  - WageIndicator Survey

- But it has the potential to become an immensely useful resource

SSHOC
social sciences & humanities open cloud

Sustainability and exploitation

- New data will become available
- For a corpus, bigger is better
- New functions will be needed
- The technology must be updated and maintained
- How do we teach students, linguists, researchers, translators to use this fabolous tool?
- How do we keep it up to date and alive?

UPF will sustain the MCSQ for 2 years

- Further investments are needed for a long term vision

SSHOC
social sciences & humanities open cloud

The new home for the MCSQ

[MCSQ]: The Multilingual Corpus of Survey Questionnaires (uib.no)

Also accessible at (http://easy.mcsq.upf.edu/)

Important now:

- Finding the resources for
  - Expansion into more studies
  - Expansion in time
  - Adding future studies
  - Developing strategies for automatic uploads and treatment of new surveys

SSHOC
social sciences & humanities open cloud

How to cite the MCSQ

Zavala-Rojas, D., Sorato, D., Hareide, L., & Hofland, K. (forthcoming 2022).Multilingual Corpus of Survey Questionnaires: a tool for refining survey translation. *Meta: Journal Des Traducteurs.*

SSHOC
social sciences & humanities open cloud

Works cited

Davidov, E., & De Beuckelaer, A. (2010). How Harmful are Survey Translations? A Test with Schwartz's Human Values Instrument. *International Journal of Public Opinion Research*, 22(4), 485–510. https://doi.org/10.1093/ijpor/edq030

Hareide, L. (2013). *The Norwegian-Spanish Parallel Corpus*. http://hdl.handle.net/11509/73

Hareide, L., & Hofland, K. (2012). Compiling a Norwegian-Spanish parallel corpus. In M. Oakes & M. Ji (Eds.), *Quantitative methods in corpus-based translation studies: A practical guide to descriptive translation research* (pp. 75–114). John Benjamins Publishing.

Harkness, Janet A. 2003. "Questionnaire Translation."In *Cross-Cultural Survey Methods*, edited by Janet A. Harkness, F. J. R. van de Vijver, and P. P. Mohler, 35–56. Hoboken: Wiley & Sons.

Harkness, J. A., Villar, A., & Edwards, B. (2010). Translation, Adaptation, and Design. In J. A. Harkness, M. Braun, B. Edwards, T. P. Johnson, L. E. Lyberg, P. P. Mohler, B.-E. Pennell, & T. W. Smith (Eds.), *Survey Methods in Multinational, Multiregional, and Multicultural Contexts* (pp. 115–140). John Wiley & Sons, Inc. https://doi.org/10.1002/9780470609927.ch7

Mohler, P. P., & Johnson, T. P. (2010). Equivalence, Comparability, and Methodological Progress. In *Survey Methods in Multinational, Multiregional, and Multicultural Contexts* (pp. 17–29). John Wiley & Sons, Inc. https://doi.org/10.1002/9780470609927.ch2

Oberski, D., Saris, W. E., & Hagenaars, J. A. P. (2007). Why are there differences in measurement quality across countries? In G. Loosveldt & M. Swyngedouw (Eds.), *Measuring Meaningful Data in Social Research*. Acco.

Zavala-Rojas, D., Saris, W. E., & Gallhofer, I. N. (2018). Preventing Differences in Translated Survey Items using the Survey Quality Predictor. In T. P. Johnson, B.-E. Pennell, I. A. L. Stoop, & B. Dorer (Eds.), *Advances in Comparative Survey Methods: Multinational, Multiregional and Multicultural Contexts (3MC)* (pp. 357–384). Wiley Series in Survey Methodology. https://doi.org/https://doi.org/10.1002/9781118884997.ch17

Zavala-Rojas, D., Sorato, D., Hareide, L., & Hofland, K. (forthcoming 2021). Multilingual Corpus of Survey Questionnaires: a tool for refining survey translation. Meta: Journal Des Traducteurs.

SSHOC

social sciences & humanities open cloud

What is the MCSQ useful for?

Quality control

- Cross check that questionnaires included in the MCSQ:

  - follow the Ask the same question (ASQ method)

  - Translations are *functionally equivalent* for the purpose of statistical analysis.

  - Concepts to be measured are kept the same same across languages

Repository of previous rounds

  - Simplify reuse of previous successful survey items

  - Root out unsuccessful or problematic survey items

Training

  - Of survey creators

  - Survey translators

  - Of automatic translation programs

Research

SSHOC

social sciences & humanities open cloud

Panel Discussion

1. In what ways can one enable interoperability of data in a service?
2. How do you decide in which format the data should be released ?
3. How did you collect the translations in the occupation database with 4000+ titles and 40+ languages?
4. What is the advantage to use the MCSQ tool to work with the questionnaires?

Advancing SSH Research with SSHOCingly good and sustainable resources

6-7 April 2022 - Brussels & Online
REGISTER NOW

Daniela Negoita
Tilburg University

Archana Bidargaddi
NSD

Fabio Franzese
Max Planck Institute
SHARE

Rodrigo Reyes
European Values Study,
Tilburg University

Lidun Hareide
Møreforsking

SSHOC
social sciences & humanities open cloud