Deliverable Number: D8.19

Deliverable Title: Training event on the surveycodings tools

Work Package: WP8 surveycodings

Deliverable type: Report

Dissemination status: Public

Submitted by: ESS ERIC

Author: Elissa Sibley (HQ/City)

Date Submitted: August 2019

Acknowledgements

Thank you to all of the SERISS colleagues and other contributors who participated in the workshop, particularly: domain experts Kea Tijdens (UvA), Silke Schneider (ESS, GESIS), Michele Belloni (SHARE, Ca’ Foscari University of Venice) and Stephanie Stuck (SHARE, MEA); technical experts Maurice Martens, Eric Balster, Iggy van der Wielen and Bart van Nieuwburg (SHARE, CentERdata); special guests Valeria Cirillo and Dario Guarascio (INAPP – Public Policy Innovation) and Alfonso Rosolia (Banca d’Italia); hosts Agar Brugiavini and Michele Belloni, and organisational team Martina Sguazzin and Marco Parussolo (SHARE, Ca’ Foscari University of Venice).

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 654221.
SERISS (Synergies for Europe’s Research Infrastructures in the Social Sciences) aims to exploit synergies, foster collaboration and develop shared standards between Europe’s social science infrastructures in order to better equip these infrastructures to play a major role in addressing Europe’s grand societal challenges and ensure that European policymaking is built on a solid base of the highest-quality socio-economic evidence.

The four year project (2015-19) is a collaboration between the three leading European Research Infrastructures in the social sciences – the European Social Survey (ESS ERIC), the Survey of Health Ageing and Retirement in Europe (SHARE ERIC) and the Consortium of European Social Science Data Archives (CESSDA AS) – and organisations representing the Generations and Gender Programme (GGP), European Values Study (EVS) and the WageIndicator Survey.

Work focuses on three key areas: Addressing key challenges for cross-national data collection, breaking down barriers between social science infrastructures and embracing the future of the social sciences.

Please cite this deliverable as: Sibley, E. (2019) *Training event on the surveycodings tools* Deliverable 8.19 of the SERISS project funded under the *European Union’s Horizon 2020 research and innovation programme* GA No: 654221. Available at: [www.seriss.eu/resources/deliverables](http://www.seriss.eu/resources/deliverables)
### Contents:

Summary ................................................................................................................................. 4  
Introduction .......................................................................................................................... 4  
Goals .................................................................................................................................. 4  
Description of the event ...................................................................................................... 5  
SERISS presentations at the workshop ................................................................................ 6  
Discussion ............................................................................................................................ 7  
  Overall response to the tools ............................................................................................ 7  
  Issues to think about .......................................................................................................... 8  
  Potential for future collaboration with respect to harmonised socio-economic variables .... 9  
  Community building ......................................................................................................... 10  
  What attendees will take from the workshop .................................................................... 10  
Next steps ............................................................................................................................ 11  
References ........................................................................................................................... 11  
Annex 1 ............................................................................................................................... 12  
Annex 2 ............................................................................................................................... 14  
Annex 3 ............................................................................................................................... 15  
  Clarifications by surveycodings experts in response to questions .................................... 15  
    Occupation (Kea Tijdens) ............................................................................................... 15  
    Education (Silke Schneider) ............................................................................................ 15  
    General/Technical (Maurice Martens) ............................................................................ 16
Summary

The SERISS surveycodings tools were showcased at an event in Venice, which brought together representatives from a diverse array of surveys and survey organisations, including the World Values Survey Association, InGRID-2, Ipsos MORI, Kantar Public, and the Leibniz Institute for Educational Trajectories. Designed as an interactive workshop, participants and special guests from two high-profile Italian organisations attended theoretical and technical presentations by the surveycodings experts, and shared insights based on their organisation’s experience. The programme also included practical sessions where participants could use the tools and meet with the experts to discuss issues relating to coding and/or using the tools. Here we give an overview of the workshop, the comments and queries raised, and our next steps for progressing the development of these tools.

Introduction

The SERISS surveycodings tools offer resource-light alternatives to manual coding for variables measuring occupation, industry, education, and field of education. The tools are available in multiple languages, and produce harmonised data to account for cross-national differences in education systems and support the coding of responses to open questions on industry and occupation (for further information see surveycodings.org).

The SERISS surveycodings workshop was designed as a dissemination and training event, to raise awareness of the surveycodings tools and to support and encourage their use. At the time of the workshop, the tools were at a stage where they needed pilot testing alongside surveys’ existing approaches to collecting data on respondents’ socio-economic characteristics, to gather evidence regarding their effectiveness and accuracy. Those who were invited to attend the workshop as participants had expressed interest in using one or more of the tools in their surveys. Communication with participants in advance of the workshop was designed to begin building a relationship between the surveycodings technical team and participants, and to ensure engagement during the event.

Goals

The SERISS surveycodings workshop sought to meet the needs of the participants as well as the Workpackage. Its formal goals were to:

1. Disseminate the surveycodings tools and encourage pilot testing.
2. Gain an overview of who is interested in the tools, how easy/difficult it is for survey holders and programmers to engage with/use them, and how they intend to do so.
3. Support participants in using the tools and make sure they feel comfortable asking for help.
4. Increase participants’ understanding of the surveycodings tools and how to use them.
5. Afford participants the opportunity to give input to the tools and their underlying databases to ensure quality.
6. Enable participants to find out what other surveys are doing and discuss the difficulties of existing coding approaches.
7. Provide an opportunity for participants to network.
8. Provide participants with support for standardised coding approaches, while minimising manual work and thus saving costs.

Description of the event

The surveycodings workshop took place at the Economics Department of Ca’ Foscari University of Venice, on 14th – 15th February 2019. The workshop was intended to raise awareness and promote the use of the surveycodings tools that were developed as part of the SERISS project. As mentioned above, the tools offer the means to automatically or semi-automatically code respondents’ occupation, industry, education, and field of education, cross-nationally and without the need for resource-intensive manual coding. For each domain it is possible to code using the live search function on the surveycodings website, or download the relevant database to integrate it into the survey’s CAWI or CAPI interface (this is the same database that underlies the live search online).

Another key goal of this workshop was to begin building a community of expert users who will contribute to the testing and maintenance of these instruments. The SERISS surveycodings team hoped the event would encourage participants to test these tools in their own surveys, alongside their existing measures, and provide feedback on their performance. In addition, they hoped to encourage comments on the content of the databases of occupations, industries, educational qualifications and fields of education that underlie the tools.

The surveycodings experts gave presentations on the main domains (occupation, industry, education, and field of education) covered by the surveycodings tools and the coding functions available for each. Technical colleagues demonstrated how to use the tools. An open session enabled participants to meet one-to-one with surveycodings colleagues. This included conversations with the domain experts, as well as an open workshop for participants to speak with technical colleagues about how to integrate the tools into their own surveys. Five participants and two special guests presented on the practices/experiences of coding socio-economic variables in their own surveys/institutions (as briefly discussed below). The workshop also provided networking opportunities between participants, special guests and SERISS colleagues involved in the surveycodings project.

In advance of the workshop, participants were asked to return a form outlining their survey’s interest in the surveycodings tools: the domain(s) they would be interested in using and in which mode(s). In response to this, a technical colleague from CentERdata arranged to speak to each participant to find out more about each individual’s role and experience, their survey’s current practices with regard to measuring and coding the domains targeted by the surveycodings tools, and what they hoped to gain from the workshop. This initial conversation served multiple purposes. In the first instance, it ensured the technical sessions would be pitched at an appropriate level for the audience and address any concerns or gaps in understanding. Secondly, it opened a dialogue that enabled technical colleagues to begin building a relationship with the participants.
Following the workshop, the SERISS surveycodings technical team will make contact with participants to answer any questions and, where appropriate, offer assistance in setting up the tools in their own surveys.

SERISS presentations at the workshop

On Day 1, presentations by SERISS colleagues provided a general introduction to surveycodings, and more detailed introductions by the domain experts.

Agar Brugiavini (Ca’ Foscari University of Venice) talked about the importance of accurate coding of socio-economic variables, giving important context to the need for the surveycodings tools at the present time.

Sarah Butt (ESS ERIC) talked about how the surveycodings project is seeking to help with the challenges of coding socio-economic variables in multinational, multicultural, multilingual surveys, and set this task in the context of the wider SERISS project.

Kea Tijdens (University of Amsterdam) spoke about the surveycodings tools for Occupation and, on behalf of Michele Belloni (Ca’ Foscari University of Venice), spoke about the surveycodings tools for Industry.

Silke Schneider (GESIS) and Maurice Martens (CentERdata) introduced the surveycodings tools for Education and Fields of Education.

On Day 2, SERISS presentations focussed on how to use the surveycodings tools, as well as discussing the on-going development of the tools and the idea of building a community of experts to contribute to this.

Maurice Martens (CentERdata) showed participants the tools that are available via the surveycodings website, and ran through their technical features.

Maurice Martens (CentERdata) spoke about the outstanding decisions regarding the on-going development and maintenance of the surveycodings tools. Maurice demonstrated a proposed method for submitting feedback on the live search interface on the surveycodings website.

A split session included a technical workshop where participants could gain advice and support from technical colleagues from CentERdata, and approach the surveycodings domain experts with questions or ideas. In addition, Maurice Martens (CentERdata) introduced the surveycodings app and sought feedback on its usability. The app, currently only available for Android devices, can be used to code respondents’ education and occupation during paper and pencil interviews (PAPI) or computer-assisted interviews (CAPI). The app is a search device; it does not store or transmit information.

During the workshop, several participants gave presentations on the work being done by their organisation in relation to coding of socio-economic variables in surveys: Olivia Ryan spoke about how Ipsos MORI codes socio-economic variables; Carina Schönmoser and colleagues talked about the work done by LifBi on educational and occupational information available in the German National Education Panel Study (NEPS); Michèle Ernst Stähli from FORS talked
about the field practices and challenges of coding occupation and education in the international surveys at FORS; *Kseniya Kizilova* spoke about the work being done by the World Values Survey Association towards developing globally-applicable scales for socioeconomic variables; and, *Ine Smits* from HIVA-KU Leuven spoke about an InGRID project that seeks to integrate national working conditions surveys in Europe. In addition, we heard presentations by special guests from two high profile Italian organisations: *Alfonso Rosolia* talked about the experience of the Bank of Italy, and *Dario Guarascio* spoke about the Italian O*Net INAPP project.

Where available, these presentations can be downloaded from the SERISS website ([https://seriss.eu/resources/sen/](https://seriss.eu/resources/sen/)).

**Discussion**

During the course of the workshop, discussions and Q&As following the presentations raised important points about the readiness of the tools, the suitability of the tools for use by participants’ organisations, and key considerations that the organisations would need to bear in mind if they were to consider using them. Some participants indicated strong interest in using the tools, and/or flagged the potential for collaboration to test or further develop them. There was also some productive discussion around community building and how to maintain the databases of occupations, educational fields, etc. that underlie the tools. The key points from these discussions are given below.

**Overall response to the tools**

Participants generally welcomed the work that has been done by the surveycodings team. There was agreement that automated coding would save time and resources, and participants could see the value of using harmonised code frames (rather than relying on national code frames). In particular, the value of harmonised cross-national code frames for studies involving migrants was flagged several times. It was also noted that, as well as drawing comparisons between countries within the same survey, researchers often want to draw comparisons across time and/or different surveys. A standardised approach to coding background variables would therefore be very helpful.

Participants were currently using a range of different tools to manage the coding of socio-economic background variables, but none of these were available/suitable for general use in the way that surveycodings aims to be. For example, Ipsos uses Ascribe but this tool is not in the public domain, and the tool used by LIfBi is currently only available in German (although translation is possible).

There was a discussion on the level of detail required for socio-economic coding and whether coding to 4-digit ISCO was necessary. It was noted that respondents will not always know to which higher level category their occupation belongs, so the best way to record occupation accurately is to collect detailed information (even if this is later aggregated; sample sizes for each 4-digit code would likely be very small). It was suggested that there could be greater emphasis on the policy relevance of having information that’s coded to such a detailed level,
and also how this information can be used in practice. Kea Tijdens’ (2016) report on the ESeG-2014 coding scheme provides the coding scheme and syntax to convert responses to the survey questions into ESeG-2014 classifications (Tijdens, 2016: 1).

A number of organisations including Ipsos, FORS and WVSA expressed interest in potentially using the tools in their surveys and in working more closely with the surveycodings team to test the tools in the field. Colleagues from LiffBi expressed particular interest in using the education tools, potentially in their 2021 survey or the survey scheduled for 2024/25. However, it was stressed that the tools would need to be in a sufficient state of readiness (for example, with databases complete for all required languages) before this would be possible. There were also a number of issues identified that could prevent uptake and which would require further investigation. These are outlined in the next section.

Issues to think about

SERISS surveycodings experts, and workshop participants, flagged the following issues to think about with respect to the potential roll out of the surveycodings tools in the field:

- For occupations (and potentially fields of education), language versions matter: more work should be done to compare different versions of the same language (e.g. German German versus Austrian German).
- In some countries, CATI is more prevalent than CAPI. Tools to be used with CATI surveys would be welcome. However, it was noted that there is often a language barrier between the interviewer and respondents who were educated overseas. For this reason, it is unlikely that the educational attainment tool would be useful in a CATI survey involving first generation migrants, unless they are asked using national education categories only.
- It will be necessary to test how survey respondents behave when using the tools for self-coding. For example, to ascertain whether respondents satisfice and pick an occupational title that looks ‘good enough’ rather than searching for the one that is most appropriate. However, it was noted that errors also occur with interviewer/office coding and there is a possibility that self-coding might be better.
- People who use the tools could create a default open answer option in their questionnaire. This would enable the retention of any open text that has been provided by respondents who cannot find an appropriate occupational title, etc.. This text could be post-coded by researchers and may help to identify errors or omissions in the underlying database. There are risks to allowing open code responses, for example, it may encourage respondents to simply type in their occupation rather than using the tool to auto-code as intended. However, this could be mitigated by retaining the terms that are entered into the search box rather than having a designated open response option. While this is in place for Education, it is down to the user to add this function for the other concepts – it is not a feature of the tool per se.
- In order to make the selection task easier, it would be useful to consider different algorithms for prioritising the codes that are shown to respondents/interviewers in response to their search terms.
- For obtaining feedback from tool users (for example, regarding omissions from the database), a feedback mechanism could be built into the database live search.
For interviewer-administered surveys, interviewers would need to be trained carefully in how to use the tools.

Before an organisation could make a final decision regarding tool uptake, they would need more information about how the surveycodings tools could be integrated with their existing software.

It was queried whether survey agencies would be willing to adopt software that had been developed by another organisation. This may be resolved by writing the requirement to use such software into any invitation to tender.

It may be difficult to ‘sell’ harmonised questions and the harmonised coding of key background variables to national survey teams. They are likely to be invested in using their own national questions and codes, and may want a simpler or more complex approach than that which is offered by the tools. For national surveys that do not involve a substantial number of first generation migrants educated abroad, it is not recommended to use the surveycodings education tool in CAPI surveys given the complexity of implementation. However, it is recommended to use national questions and response options that allow mapping to the detailed 3-digit ISCED 2011 scheme.

Concerns were raised about the long-term sustainability of the tools and how databases and software would be maintained after the end of the SERISS project.

Potential for future collaboration with respect to harmonised socio-economic variables

In addition to participants expressing potential willingness to adopt or pilot test the surveycodings tools in their own surveys (subject to them being in a suitable state of readiness), the following areas were identified for possible collaboration in the field of harmonised socio-economic coding:

- There are some links with the work of InGRID-2 and there may be possibilities to collaborate. The InGRID team offered to raise awareness of the surveycodings tools with their General Assembly and other stakeholders.
- The special guest from the Bank of Italy noted that the tools could be applied to firm surveys, to collect firm data to the same standard as household data and fill gaps in the information that’s currently available. It would be possible to code for the following:
  - occupation and the skills used in different jobs, for example to enable the measurement of labour demand;
  - the cognitive requirements of a set of jobs, to identify the jobs that people are capable of doing during their working life;
  - the products a firm generates, and the function of the firm;
  - which roles are associated with social exclusion.

INAPP is working on data collection in this direction. Firm data could also be used to look at health, and compare outcomes for people who do the same jobs in different regions.
Community building

Particular attention was paid to the possibility of building a community of experts to quality assure and update the databases on occupation, industry and education that underlie the surveycodings tools.

One suggestion was to put out a call for reviewers; for example, to systematically review the whole Search Tree for the country/countries they know well. This could be particularly helpful for checking whether the Education classifications are structured correctly for countries that the surveycodings team do not know well enough. This could involve developing a review panel.

The surveycodings experts flagged that they would particularly value input on a number of points, including:

- Assistance finding information about ISCED Fields of Education 2013 in Italian at 2 and 3 digit levels.
- Suggestions for algorithms for use in the live search, that might work better than those currently in use.
- Comments on the idea of having a feedback mechanism built into the live search interface on the website, for users to submit comments regarding omissions or inaccuracies.
- Ideas for the best way of organising a workflow to ensure the databases remain up-to-date, are quality assured and can bring in input from data users. Ideas put forward were: users to be notified of database updates via version notes; a newsletter that includes a summary of database updates and links to the relevant version notes; integrating the feedback mechanism further into the backend of the tools, so that users can link their feedback to the relevant classification/code. The idea of a forum was ruled out, due to the difficulty of maintaining and sustaining this type of communication platform.

The surveycodings team would also welcome expressions of interest from those who would like to be on a review panel for occupation or education. In addition, suggestions would be welcome regarding the types of people it might be appropriate to invite and how they might be recruited. One idea for how to recruit reviewers, was to add a 'subscribe' button to the website; having clicked on this button, the user would be presented with a short series of questions that ask whether they would be interested in being on a review panel and, if so, for which domain(s), country/countries and language(s). The following were also suggested: asking employment agencies to recommend ways to find occupation experts within their country; and, publishing the lists of occupations and education classifications in each country, and asking people to review these lists (similarly to how it is implemented for SHARE). It was flagged that very clear guidance would be needed for anyone who is asked to review the lists, as well as information on why it is being done; otherwise reviewers might use national coding and different logics.

What attendees will take from the workshop

In discussing what they will take from the workshop, attendees mentioned points in several key areas: an understanding of the complexity of the surveycodings tools and why harmonised
coding is important; a clearer understanding of how survey coding fits into the bigger picture; knowledge of the activities of others working in this field; awareness of the potential to work collaboratively with the surveycodings experts on testing and further developing the tools; and, seeing a need to reconsider their organisation’s approach to survey coding, and/or think about whether their organisation could use the surveycodings tools.

Next steps

Following the workshop, the surveycodings team will continue to work on the tools and ensure that they are in the best state of readiness possible at the end of the SERISS project in August 2019. A project update will be sent out to workshop participants in August 2019 summarising for each tool and underlying database:

- the current status of the tool;
- any omissions/known bugs;
- any plans for further work (along with a timetable for doing so);
- details of who to contact for more information.

Training materials will be developed for the surveycodings tools and made available via the surveycodings website. Kristi Winters, who leads on training activities for SERISS, will be in contact with workshop participants to gain a better understanding of which training materials would be most helpful.

Further thought will be given to how to build a surveycodings user community, and to the crucial issues relating to updating and versioning the databases that underlie the tools.

Potential collaborations with respect to piloting the tools and comparing the performance of the automated coding tools/harmonised databases with manual coding/national code frames will be followed up.

Surveycodings colleagues will follow up with InGRID colleagues regarding possible cross-fertilisation between the two teams’ strands of work: work on the surveycodings tools, and the work done by InGRID on understanding the importance and effects of socio-economic characteristics.

References


SERISS WP8 surveycodings workshop

Room Partesotti (4th floor), Economics Campus San Giobbe
Department of Economics
Ca’ Foscari University of Venice

14-15 February 2019

Agenda

Thursday 14th February, 12:00 – 19:00 (Dinner at 19:30)

12:00 Lunch
14:00 Welcome (Agar Brugiavini, Ca’ Foscari University of Venice)
14:05 Tour de table (Agar Brugiavini, Ca’ Foscari University of Venice)
14:20 Introducing surveycodings (Agar Brugiavini, Ca’ Foscari University of Venice and Sarah Butt, ESS ERIC)
14:50 Introduction to the surveycodings domains and tools
   - Occupation and industry (Kea Tijdens, University of Amsterdam)
   - Education (Silke Schneider, GESIS and Maurice Martens, CentERdata)
15:50 Tea/coffee
16:05 Participant presentations (Chair: Stephanie Stuck, SHARE MEA-MPISOC)
   - Coding Socio-economic variables at Ipsos (Olivia Ryan, Ipsos MORI)
   - Educational and Occupational Information in the German National Education Panel Study (NEPS) (Carina Schönmoser, LIfBi)
   - Coding occupation and education in the international surveys at FORS (Switzerland): field practices and challenges (Michèle Ernst Stähli, FORS)
17:05 Special Guest presentation
   The Experience of the Bank of Italy (Alfonso Rosolia, Banca d’Italia)
17:25 Tea/coffee
17:40 Participant presentations (Chair: Stephanie Stuck, SHARE MEA-MPISOC)
   - Working towards the globally-applicable scales for socioeconomic variables: the World Values Survey perspective and experience (Kseniya Kizilova, WVSA)
   - Integrating national working conditions surveys in Europe: a test (Ine Smits, HIVA-KU Leuven)
   - (TBC)
18:40 Wrap-up and Q&A (Chair: Sarah Butt, ESS ERIC)
18:55 Closing Day 1 (Michele Belloni, Ca’ Foscari University of Venice)
19:30 Dinner at Laguna Libre (http://www.lagunalibre.it/it/)

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 654221.
Friday 15th February, 08:30 – 14:00

08:30  Tea/coffee
09:00  Welcome back (Sarah Butt, ESS ERIC)
09:10  How to use the surveycodings tools (Maurice Martens, CentERdata)
10:30  Tea/coffee
10:45  Split session:
       • Technical workshop (Eric Balster, Maurice Martens, Iggy van der Wielen and Bart van Nieuwburg, CentERdata)
       • Ask a domain expert (Michele Belloni, Ca’ Foscari University of Venice; Silke Schneider, GESIS; Stephanie Stuck, SHARE MEA-MPISOC; Kea Tijdens, University of Amsterdam)
11:35  On-going development of the surveycodings tools and community-building (Maurice Martens, CentERdata)
12:35  Special Guest presentation
       Occupations, skills and duties: The Italian O*Net INAPP project (Dario Guarascio, INAPP - Public Policy Innovation)
12:55  Closing Day 2 (Michele Belloni, Ca’ Foscari University of Venice and Sarah Butt, ESS ERIC)
13:00  Lunch (opportunity for continued technical workshop and discussions with experts)
14:00  End
**SERISS WP8 surveycodings workshop**

Room Partesotti (4th floor), Economics Campus San Giobbe  
Department of Economics  
Ca’ Foscari University of Venice  

14-15 February 2019

List of Participants

<table>
<thead>
<tr>
<th>Name</th>
<th>Organisation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eric Balster</td>
<td>CentERdata – Institute for Data Collection and Research</td>
</tr>
<tr>
<td>Michele Belloni</td>
<td>Ca’ Foscari University of Venice</td>
</tr>
<tr>
<td>Janna Besamusca</td>
<td>InGRID2 (supporting expertise in inclusive growth)</td>
</tr>
<tr>
<td>Agar Brugiavini</td>
<td>Ca’ Foscari University of Venice</td>
</tr>
<tr>
<td>Sarah Butt</td>
<td>ESS ERIC</td>
</tr>
<tr>
<td>Danilo Cavapozzi</td>
<td>Ca’ Foscari University of Venice</td>
</tr>
<tr>
<td>Valeria Cirillo</td>
<td>INAPP – Public Policy Innovation</td>
</tr>
<tr>
<td>Michèle Ersnt Stähli</td>
<td>FORS (Swiss Centre of Expertise in the Social Sciences)</td>
</tr>
<tr>
<td>Dario Guarascio</td>
<td>INAPP – Public Policy Innovation</td>
</tr>
<tr>
<td>Kseniya Kizilova</td>
<td>World Values Survey Association (WVSA)</td>
</tr>
<tr>
<td>Miquelle Marchand</td>
<td>CentERdata – Institute for Data Collection and Research</td>
</tr>
<tr>
<td>Maurice Martens</td>
<td>CentERdata – Institute for Data Collection and Research</td>
</tr>
<tr>
<td>Sven Pelz</td>
<td>Leibniz Institute for Educational Trajectories (LiBi)</td>
</tr>
<tr>
<td>Alfonso Rosolia</td>
<td>Banca d’Italia</td>
</tr>
<tr>
<td>Olivia Ryan</td>
<td>Ipsos MORI</td>
</tr>
<tr>
<td>Silke Schneider</td>
<td>GESIS – Leibniz Institute for the Social Sciences</td>
</tr>
<tr>
<td>Carina Schönmoser</td>
<td>Leibniz Institute for Educational Trajectories (LiBi)</td>
</tr>
<tr>
<td>Elissa Sibley</td>
<td>ESS ERIC</td>
</tr>
<tr>
<td>Ine Smits</td>
<td>HIVA-KU Leuven</td>
</tr>
<tr>
<td>Stephanie Stuck</td>
<td>SHARE MEA-MPISOC</td>
</tr>
<tr>
<td>Sam Sullivan</td>
<td>Kantar Public UK</td>
</tr>
<tr>
<td>Kea Tijdens</td>
<td>University of Amsterdam</td>
</tr>
<tr>
<td>Iggy van der Wielen</td>
<td>CentERdata – Institute for Data Collection and Research</td>
</tr>
<tr>
<td>Bart van Nieuwburg</td>
<td>CentERdata – Institute for Data Collection and Research</td>
</tr>
<tr>
<td>Tina Vovk</td>
<td>Faculty of Social Sciences, University of Ljubljana</td>
</tr>
<tr>
<td>Chris White</td>
<td>European Foundation for the Improvement of Living and Working Conditions (Eurofound)</td>
</tr>
<tr>
<td>Kristi Winters</td>
<td>GESIS – Leibniz Institute for the Social Sciences</td>
</tr>
<tr>
<td>Špela Zajšek</td>
<td>Faculty of Social Sciences, University of Ljubljana</td>
</tr>
<tr>
<td>Markus Zielonka</td>
<td>Leibniz Institute for Educational Trajectories (LiBi)</td>
</tr>
</tbody>
</table>

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 654221.
Annex 3

Clarifications by survey codings experts in response to questions

The SERISS survey codings experts provided a number of clarifications regarding the tools, in response to questions from participants. The experts were: Kea Tijdens (Occupation coding tools), Silke Schneider (Education and Fields of Education coding tools), and Maurice Martens (technical issues regarding the tools and the website).

Occupation (Kea Tijdens)

*How the database is constructed*

- Where two occupations are translated in the same way, we retained the lowest-skilled occupation. A pre-defined set of instructions advised translators regarding which occupations are higher-skilled / lower-skilled (see Deliverable D8.7, Appendix 1 Instructions for translators for further information).
- Duplicate translations were flagged automatically; it was not necessary to look for them manually.
- The ISCO manual is rarely translated. However, we compared our coding manual to those created by statistical offices, comparing 16 coding indexes (coding different occupation titles) that had at least 1000 entries.
- The rules for translation were not compared with those in the ISCO coding manual (for example, how to treat special cases). This is one reason to build a community around the tools; to check the lists, etc..
- The identification of occupational skill level is difficult. Some occupations were upgraded to a higher level between ISCO 88 and ISCO 08, with consideration to changes in the cognitive demands of the occupation. However, there is no measure for the skill required for upgrade. There is a lot of work still to be done in this area.

Education (Silke Schneider)

*How the database is constructed*

- To cope with within-country changes in the education system over time, we took information from the ISCED mappings as well as looking at the measures used in surveys of adults. We then classified them according to ISCED definitions. Where necessary, we spoke with colleagues, experts, etc. in other countries to find out more detail for the classification.
- Regarding outdated qualifications: there is a field for recording for which years each outdated qualification existed. However, this is not complete and in some cases we have had to specify an approximate end date (for example, a decade) because sometimes qualifications are phased out over a number of years. This may be completed in the future if funding becomes available.
- It has not been possible to set up the tool to cross-reference the respondent’s date of birth and show only those qualifications that should be relevant to them. This would require integrating information on respondent age in the search algorithm and
matching this with the as yet incomplete information on which qualifications existed when. The algorithm is already quite complex so this was deferred.

- There is interest in building a similar database for parents’ education, but it is important to bear in mind that asking respondents to tell us their parents’ educational qualifications is challenging; respondents will often not know this in very much detail. For this reason, the surveycodings tool does not appear to be useful for proxy reporting of education.

**Functionality of the tool**

- It cannot be guaranteed at this stage that all outdated qualifications are included.
- The Education database cannot yet derive years of education. We would like to add this when funding becomes available.
- The education coding tool has not been designed for telephone interviews because it requires the visual communication channel, especially when asking first generation migrants on potential foreign qualifications. These tools are designed with migrant surveys in mind, and those surveys rely on respondents reading the items themselves because the interviewer does not speak the respondent’s native language (educational qualifications in the database use the language of the respective educational system since in most cases these cannot be translated). Obviously, this would not work over the phone. In a general population survey, the interviewer could sit at a computer, read out the questions to the respondent over the phone, then type in their answers and read out the responses suggested by the tool. But the tool hasn’t yet been set up with that in mind (for example, the lists of qualifications might be too long to make this practical). It also might not be possible to implement the codes developed for the CAPI systems in CATI systems. We are bearing this in mind to consider as a potential future development if funding becomes available.
- For the database live search: we are still working on the algorithm that will be used when two or more words are entered into the text-search.

**General/Technical (Maurice Martens)**

- If an organisation wants to use a tool in their own CAPI survey, it should be possible to set this up using the resources that are available on the website. However, CentERdata would be happy to help as the APIs needed will vary depending on the CAPI software. It may make sense, for popular CAPI software, for CentERdata to set that up for one party and then redistribute it where needed.
- During a live search of the database, it is possible to search using just text-matching. However, to obtain a more succinct list of possible matches it would be possible to set up the site using other algorithms.
- In the Education database, results are sorted by ISCED level (according to the international order, rather than an order specific to that country). Also, a hit that matches the beginning of the word used in the text-search will be listed above those that match the middle or end of the word.
- The algorithm used for the live search does not have a negative impact on the running time.
- Dependent interviewing and integration of preloaded data is possible for the surveycodings modules, if the module were adapted for this purpose. This is not
currently part of the functionality of the tools, but it could be developed. There are technical challenges around developing this, but interested users could talk to CentERdata.

- During data collection, depending on the type of survey it is possible to use the tools in different ways. This depends largely on the respondent or interviewer’s device. If using a CAPI survey with no internet connection, the database needs to be already downloaded and would be distributed by the survey agency on each CAPI device. For a CAWI survey, the relevant database does not need to be downloaded as it can connect to the online service. If using the app, it is possible to connect to the server to download the domain/language set that is relevant to the interview. This would require an internet connection, but after that the set is stored in the app and the downloaded domains can be accessed without an internet connection.

- It should be possible for the Stata code that automatically sets certain classifications for Education, to be used also for Occupation to automatically derive ISEI codes (rather than the user having to do this separately).